Mnews: A Study of Multilingual News Search Interfaces

Chenjun Ling

Follow this and additional works at: https://scholarcommons.scu.edu/eng_phd_theses

Part of the Computer Engineering Commons
I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Chenjun Ling

ENTITLED

Mnews: A Study of Multilingual News Search Interfaces

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE AND ENGINEERING
MNEWS: A STUDY OF MULTILINGUAL NEWS SEARCH INTERFACES

by

Chenjun Ling

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Computer Science and Engineering
School of Engineering
Santa Clara University

Santa Clara, California

September 2019
Abstract

With the global expansion of the Internet and the World Wide Web, users are becoming increasingly diverse, particularly in terms of languages. In fact, the number of polyglot Web users across the globe has increased dramatically.

However, even such multilingual users often continue to suffer from unbalanced and fragmented news information, as traditional news access systems seldom allow users to simultaneously search for and/or compare news in different languages, even though prior research results have shown that multilingual users make significant use of each of their languages when searching for information online.

Relatively little human-centered research has been conducted to better understand and support multilingual user abilities and preferences. In particular, in the fields of cross-language and multilingual search, the majority of research has focused primarily on improving retrieval and translation accuracy, while paying comparably less attention to multilingual user interaction aspects.

The research presented in this thesis provides the first large-scale investigations of multilingual news consumption and querying/search result selection behaviors, as well as a detailed comparative analysis of polyglots’ preferences and behaviors with respect to different multilingual news search interfaces on desktop and mobile platforms. Through a set of 4 phases of user studies, including surveys, interviews, as well as task-based user studies using crowdsourcing and laboratory experiments, this thesis presents the first human-centered studies in multilingual news access, aiming to drive the development of personalized multilingual news access systems to better support each individual user.
Keywords

Multilingual Search; Multilingual Interfaces; Human–Computer Information Retrieval; User Studies; Crowdsourcing; Desktop; Mobile; Eyetracking.
Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisors, Dr. Ben Steichen and Dr. Silvia M. Figueira. My journey towards the successful completion of the Ph.D. program would not have been possible without the support from both of them. I learned a lot from them, and I have been very lucky and grateful to have them as my advisors during this wonderful journey of Ph.D. study.

I particularly would like to express my appreciation to Dr. Ben Steichen for bringing me into the exciting research field of multilingual news and providing me with the great guidance and encouragement throughout the whole journey. He encouraged me to pursue the research direction that I got very interested in, and patiently provided me with his advice whenever I was in need of guidance. The experience of working with him has been a great pleasure and rewarding, his guidance and advice have been critical for my growth into an independent researcher.

I also particularly would like to express my appreciation to Dr. Silvia M. Figueira for accepting me under her supervision after Dr. Steichen left SCU and supporting me to continue pursuing the Ph.D. studies at SCU. Her belief in me, insightful advice for my research and paper writing, and encouragement have played an important role for me to complete my research studies.

I would like to thank all the members of my thesis committee, who are Dr. Yi Fang, Dr. Maya Ackerman, and Dr. Ahmed Amer, for their continuing support, encouragement, and advice for my research studies. In addition, I would like to thank all the professors who have given me valuable knowledge in their courses and advice during my Ph.D. studies at SCU.

I would like to express my heartfelt gratitude for the funding support that Packard Engineering Research Fellowship had provided me through my first year Ph.D. study. I would like to thank the participants of my research user studies, my research work would not be possible without the participation and feedbacks of them. I would like to thank Tiffany Lee for helping recruit some participants and run some of the user studies in phase
3. I also would like to thank Jiaoni Zhou and Xiaoyu Li for helping run part of phase 1 user study. I would like to thank Alexander Choulos for his help during the phase 1 prototype development and data analysis.

Last but not the least, I would like to thank my husband and my parents, for their unconditional love, encouragement, and most importantly for believing in me.
# Table of Contents

Abstract .................................................................................................................. ii

Keywords ............................................................................................................... iii

Acknowledgments ................................................................................................. iv

Table of Contents ................................................................................................. vi

List of Figures ........................................................................................................ x

List of Tables .......................................................................................................... xiv

Chapter 1 Introduction .......................................................................................... 1

1.1 Introduction ...................................................................................................... 1

1.2 Objectives and Research Questions .................................................................. 2

1.2.1 Phase 1: Multilingual search engine preferences and behaviors
& news consumption behavior pilot ................................................................. 2

1.2.2 Phase 2: Polyglot news access behaviors with respect to
different media platforms, topics, and device modalities ......................... 3

1.2.3 Phase 3: Designing and evaluating different multilingual news
search interface designs for mobile devices ................................................. 3

1.2.4 Phase 4: Comparing user preferences and behaviors between
desktop and mobile multilingual news search interfaces ......................... 4

1.3 Methodology ................................................................................................... 4

1.3.1 Phase 1: Multilingual Search Engine User Interface Study &
Pilot Multilingual News Consumption Behavior Interview ....................... 4

1.3.2 Phase 2: Survey and Task-Based User Study (Through
Crowdsourcing) ............................................................................................... 5

1.3.3 Phase 3: Lab-Controlled Interview ............................................................... 5

1.3.4 Phase 4: Task-Based User Study (Through Laboratory
Experiment) .................................................................................................... 6
### Contribution

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>6</td>
</tr>
</tbody>
</table>

### Thesis Overview

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>10</td>
</tr>
</tbody>
</table>

### Chapter 2: Related Work

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>10</td>
</tr>
</tbody>
</table>

### Chapter 3: Phase 1: Multilingual Search Engine Preferences and Behaviors & News Consumption Behavior Pilot

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>12</td>
</tr>
<tr>
<td>3.2</td>
<td>30</td>
</tr>
</tbody>
</table>

#### 3.1 Phase 1.1 A Comparative User Study of Interactive Multilingual Search Interfaces

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1</td>
<td>12</td>
</tr>
<tr>
<td>3.1.2</td>
<td>13</td>
</tr>
<tr>
<td>3.1.3</td>
<td>18</td>
</tr>
<tr>
<td>3.1.4</td>
<td>26</td>
</tr>
</tbody>
</table>

#### 3.1.1 Experimental Setup

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1.1</td>
<td>13</td>
</tr>
</tbody>
</table>

#### 3.1.2 Results

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.2.1</td>
<td>18</td>
</tr>
</tbody>
</table>

#### 3.1.3 Summary and Discussion

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3.1</td>
<td>26</td>
</tr>
</tbody>
</table>

#### 3.1.4 Conclusion

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.4.1</td>
<td>29</td>
</tr>
</tbody>
</table>

#### 3.2 Phase 1.2 Pilot Multilingual News Consumption Behavior Interview

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1</td>
<td>30</td>
</tr>
<tr>
<td>3.2.2</td>
<td>30</td>
</tr>
<tr>
<td>3.2.3</td>
<td>31</td>
</tr>
<tr>
<td>3.2.4</td>
<td>34</td>
</tr>
</tbody>
</table>

#### 3.2.1 Introduction and Research Questions

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1.1</td>
<td>30</td>
</tr>
</tbody>
</table>

#### 3.2.2 Experiment Setup

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.2.1</td>
<td>30</td>
</tr>
</tbody>
</table>

#### 3.2.3 Study Results

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.3.1</td>
<td>31</td>
</tr>
</tbody>
</table>

#### 3.2.4 Summary and Current Impediments

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.4.1</td>
<td>34</td>
</tr>
</tbody>
</table>

### Chapter 4: Phase 2: An Investigation of Consumption, Querying, and Search Result Selection Behaviors

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>36</td>
</tr>
<tr>
<td>4.2</td>
<td>37</td>
</tr>
</tbody>
</table>

#### 4.2 Methodology

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1</td>
<td>37</td>
</tr>
<tr>
<td>4.2.2</td>
<td>37</td>
</tr>
</tbody>
</table>

#### 4.2.1 Phase 2.1: Multilingual News Consumption Behavior Survey

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1.1</td>
<td>37</td>
</tr>
</tbody>
</table>

#### 4.2.2 Phase 2.2: Task-Based User Studies

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.2.1</td>
<td>37</td>
</tr>
</tbody>
</table>

#### 4.2.3 Participant Recruitment, Quality Control, and Payment

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.3.1</td>
<td>44</td>
</tr>
</tbody>
</table>

#### 4.2.4 News Topics, Translations, & Analysis

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.4.1</td>
<td>44</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>45</td>
</tr>
</tbody>
</table>

#### 4.3.1 Phase 2.1: Consumption Survey

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1.1</td>
<td>45</td>
</tr>
</tbody>
</table>

#### 4.3.2 Phase 2.2: Task-based User Studies

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.2.1</td>
<td>51</td>
</tr>
</tbody>
</table>
4.4 Summary and Discussion ................................................................. 60
4.5 Conclusion ..................................................................................... 64

Chapter 5 Phase 3: Mobile Pilot Study ............................................. 66

5.1 Introduction ................................................................................... 66
5.2 Experimental Setup ..................................................................... 66
5.2.1 Interface Used in the Study ....................................................... 66
5.2.2 Platform, Focus, and Participants ............................................ 68
5.2.3 Task Questions ......................................................................... 69
5.2.4 Procedure ................................................................................ 70
5.2.5 Data Analysis ......................................................................... 71
5.3 Results ......................................................................................... 71
5.3.1 Participant Demographics ...................................................... 71
5.3.2 User Preferences ...................................................................... 72
5.3.3 General User Behavior .............................................................. 76
5.4 Summary and Discussion ............................................................. 77

Chapter 6 Phase 4: Multilingual News Desktop and Mobile Comparison User Study ................................................................. 79

6.1 Introduction ................................................................................... 79
6.2 Experimental Setup ..................................................................... 80
6.2.1 Interface Used in the Study ....................................................... 80
6.2.2 Platform and Participants ....................................................... 84
6.2.3 Task Questions ......................................................................... 84
6.2.4 Procedure ................................................................................ 84
6.2.5 Data Analysis ......................................................................... 87
6.3 Results ......................................................................................... 87
6.3.1 Participant Demographics ...................................................... 87
6.3.2 Results from Desktop Interface Comparison ......................... 87
6.3.3 Results from Mobile Interface Comparison ......................... 95
6.3.4 Comparison Between Desktop and Mobile Interfaces ............... 100
6.4 Summary and Discussion ................................................................. 106

Chapter 7 Conclusion and Future Work ............................................. 111

Reference list ..................................................................................... 115
List of Figures

Figure 1 Tabbed Interface .................................................................14
Figure 2 Panel Interface .................................................................14
Figure 3 Interleaved Interface .........................................................15
Figure 4 Universal Interface ..........................................................15
Figure 5 Post-study questionnaire results. ‘Panels’ is considered the ‘most preferred’ and ‘easiest to use’ by most participants, while ‘Interleaved’ is the ‘most disliked’ .................................................................19
Figure 6 Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree) .................................................................20
Figure 7 Average query number per language (Mix indicates the use of multiple languages within a single query) ..............................................................................................................21
Figure 8 Average task time per interface ..........................................23
Figure 9 Eye tracking features per interface ......................................24
Figure 10 First glance heat maps ....................................................25
Figure 11 Average query number per language per task type ..........26
Figure 12 Task-based Study 1 – writing a query in response to a topic description .................40
Figure 13 Task-based Study 2 – choosing a query from a predefined list containing one query in L1 and one query in L2 .................................................................41
Figure 14 Task-based Study 3 – choosing between an L1 list and an L2 list .................42
Figure 15 Task-based Study 4 – choosing between an L1 list and a Mix-L1&L2 list .....43
Figure 16 Participant gender percentage per language .....................45
Figure 17 Participant age percentage per language ..........................46
Figure 18 Consumption survey platform time percentage .................47
Figure 19 Consumption survey platform language percentage ..........47
Figure 20 Consumption survey news topics language percentage

Figure 21 “Most frequent language” for participants who had previously indicated that they use “More than one language”

Figure 22 Answers to Q10. “Have you experienced any issues when consuming news in different languages,?” and Q11. “Do you feel there is a lack in news availability in your first language suggestions?”

Figure 23 Language choices for Task-based studies 1-4

Figure 24 Task-based study 2 – most frequent language choice reasons overall

Figure 25 Task-based study 2 – relative language selection reasons for English L1 vs. Non-English L1 participants

Figure 26 Task-based study 3 – most frequent language choice reasons overall

Figure 27 Task-based study 3 – relative language selection reasons for English L1 vs. Non-English L1 participants

Figure 28 Task-based study 4 – most frequent language choice reasons overall, when choosing either an L1-only result list or a Mix-L1&L2 result list

Figure 29 Task-based study 4 – most frequent language choice reasons for English L1 and Non-English L1 participants

Figure 30 Query/Result language choices depending on user L2 proficiency

Figure 31 Query/Result language choices depending on news topic/domain

Figure 32 CollapseAll Interface

Figure 33 CollapsePartial Interface

Figure 34 Slide Interface

Figure 35 Tabbed Interface

Figure 36 Overview of participants’ L1 and L2

Figure 37 Overview of participants’ L2 proficiency levels (5=native or bilingual proficiency, 1 = elementary proficiency)
Figure 38 Post-study questionnaire results ................................................................. 73

Figure 39 Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree) ................................................................. 76

Figure 40 Interface and Participant Per Interface Average Time ................................... 77

Figure 41 Desktop Tabbed Interface ........................................................................ 81

Figure 42 Desktop Panels Interface ........................................................................ 82

Figure 43 Desktop Universal Interface .................................................................... 82

Figure 44 Mobile CollapseAll Interface .................................................................. 83

Figure 45 Mobile Tabbed Interface ........................................................................ 83

Figure 46 Participants L1 and L2 ............................................................................ 87

Figure 47 Post-study questionnaire results. Desktop ‘panel’ is considered the ‘most preferred’ and ‘easiest to use’ by most participants, while desktop ‘tabbed’ is considered the ‘most disliked’ ................................................................. 88

Figure 48 Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree) ................................................................. 91

Figure 49 Participant behavior features per interface ................................................. 92

Figure 50 Participant L2 proficiency and average query number ............................ 92

Figure 51 Participant behavior features per news category ...................................... 94

Figure 52 Post-study questionnaire results. ‘collapseAll’ is considered ‘more preferred’ and ‘easier to use’ by most participants, while ‘tabbed’ is considered ‘more disliked’ .............................................................................................................. 95

Figure 53 Mid-study questionnaire results of mobile interfaces. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree) ................................................................. 97

Figure 54 Participant behavior features per mobile interface .................................. 98

Figure 55 Participant behavior features per news category .................................... 99

Figure 56 Participant platform preferences ............................................................. 101
Figure 57 Participant interface preferences: Desktop ‘Tabbed’ and Mobile ‘Tabbed’....102

Figure 58 Participant interface preferences: Desktop ‘Universal’ and Mobile ‘CollapseAll’ ................................................................. 103

Figure 59 Participant interface preferences: Desktop ‘Tabbed’ and Desktop ‘Universal’ .................................................................................. 104

Figure 60 Participant interface preferences................................................................................................................................. 105

Figure 61 Participant behavior features: Desktop and Mobile........................................................................................................ 105
List of Tables

Table 1  Task Questions and Question Types ..................................................17
Table 2  Eye-Tracking Features.................................................................18
Table 3  Query language changes within tasks ...........................................22
Table 4  Check News Methods.................................................................32
Table 5  News Categories and The Read Behavior .....................................34
Table 6  Phase 1 Survey Questions..............................................................38
Table 7  Participant Language Distributions for the Task-Based Studies .......52
Table 8  Study 2 Language Selection Reasons ...........................................55
Table 9  Study 3 Language Selection Reasons ...........................................57
Table 10 Study 4 Language Selection Reasons ...........................................59
Table 11 Phase 3 News Categories and News Topics ...................................69
Table 12 Phase 4 News Categories and News Topics ...................................85
Chapter 1

Introduction

1.1 Introduction

Since the advent of terrestrial radio, satellites, and the Internet, the world has become connected more than ever. In fact, nearly 46.1% of the world’s population is connected to the Internet today\(^1\). As a result of this global expansion, user diversity has also increased, especially in terms of user languages. In particular, in addition to the rising diversity of online users’ native languages, there are also many people across the globe now who are polyglots, i.e. people who are proficient in more than one language. For example, 21% of US residents speak a language other than English at home (with the majority also speaking English “well” or “very well”), and more than 44% of California residents speak a language other than English at home (again, with the majority also speaking English “well” or “very well”)\(^2\). Even more strikingly, in the European Union, an average of 94.6% of secondary education pupils learn English in general programs, and an average of 64.7% even learn two or more languages\(^3\). Similarly, a recent survey in China has shown that out of those with junior secondary education qualifications or above, 67.4% of Chinese graduates had studied at least one foreign language, among which as many as 93.8% had studied English [1,2]. Overall, it is estimated that half of the world’s population learns or speaks at least two languages [3].

Despite these latest statistics on Internet penetration and multiple language proficiencies, there remain many obstacles when accessing global news information. For example, information is often not available in certain languages, which may stem from the fact that some languages are vastly underrepresented in terms of Web content, despite large user populations. For example, statistics by Web Technology Surveys\(^4\) indicate that 54.1% of all current websites are written in English, whereas Chinese language websites only account for 1.7% of all current websites, even though 52.2% of online users are from China\(^5\). Likewise,

---

\(^1\)www.internetlivestats.com/internet-users/
\(^2\)https://factfinder.census.gov
\(^3\)Eurostat Foreign language learning statistics: ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_learning_statistics
\(^4\)w3techs.com/technologies/overview/content_language/all
multiple language proficiencies are typically not supported, since traditional news systems seldom allow users to simultaneously search and/or compare news in different languages. At the same time, given the many ways and platforms for news access available today (including traditional outlets as well as social media), users have to spend significant effort to check and aggregate information from different information sources, as well as additional effort to summarize fragmented information across languages.

To address these challenges, the research presented in this thesis aims to drive the development of novel multilingual news access systems and interfaces that can support the discovery and browsing of news content in multiple languages, and according to each individual user’s personal abilities (e.g. language proficiencies), preferences (e.g. language preferences, interface preferences, device preferences), and context (e.g. news topic).

1.2 Objectives and Research Questions

The objective of this research is to investigate multilingual news access behaviors and preferences, as well as to design, develop, and evaluate novel interactive multilingual news search interfaces. This research provides a better understanding of how to best support polyglot users and their multiple language skills, and provides guidelines for multilingual news access system design. To achieve these objectives, the research presented in this thesis consists of 4 phases of user-centered studies, as outlined below.

1.2.1 Phase 1: Multilingual search engine preferences and behaviors & news consumption behavior pilot.

Phase 1.1: Multilingual Search Engine User Interface Study: The objective of this study is to evaluate different interactive multilingual search interface designs, to get a better understanding of how to best support polyglot Web users and their multiple language skills during general Web search. Specifically, we ask the following research questions: (RQ1.1.1) What type(s) of interactive multilingual search interface designs are preferred by polyglot users? (RQ1.1.2) What are the interaction behaviors of polyglot users with interactive multilingual search systems? (RQ1.1.3) Do different interactive multilingual search interfaces and/or task types lead to different user behaviors?

Phase 1.2: Pilot Multilingual News Consumption Behavior Interview: To gain a better understanding of multilingual news access, this study consists of interviews with multilingual users, which helps provide insights on news browsing habits, preferences in language when browsing news, and technical difficulties encountered when browsing in a non-English
language. With the plethora of news platforms available to users, and the varying levels of multilingual information availability and preferences, this interview investigates: (RQ1.2.1) Which platforms do multilingual users typically choose to search/browse news? (RQ1.2.2.) Which topics do multilingual users typically choose to search/browse news? (RQ1.2.3) Which languages do multilingual users prefer with different news platforms and different news topics? (RQ1.2.4) How easy is it currently to receive news from a platform in different languages, and what are current barriers to (and suggestions for improving) multilingual information access?

1.2.2 Phase 2: Polyglot news access behaviors with respect to different media platforms, topics, and device modalities.

This phase aims to gather initial large-scale qualitative and quantitative data from crowdsourcing platforms regarding current user behaviors, user preferences, and access impediments, in order to inform the designs and evaluations performed in phases 3 and 4. The two main steps in this phase include:

Phase 2.1: A survey that elicits current news consumption behaviors: There are three main research questions as part of phase 2.1: (RQ2.1.1) Which platforms do multilingual users typically choose to search/browse news? (RQ2.1.2) Which topics do multilingual users typically choose to search/browse news? (RQ2.1.3) Which languages do multilingual users prefer with different news platforms and different news topics? (RQ2.1.4) How easy is it currently to receive news from a platform in different languages and what are current barriers (and suggestions for improvement) to multilingual information access?

Phase 2.2: A large-scale user study that elicits user querying and result selection behaviors: there are 3 main research questions: (RQ2.2.1) Which of their language(s) do multilingual users generally use (and why) when querying and choosing search results? (RQ2.2.2) What is the effect of language proficiency on multilingual user query constructing/selection or result selection behavior? (RQ2.2.3) What are the effects of news domain/topic on multilingual user result selection behavior?

1.2.3 Phase 3: Designing and evaluating different multilingual news search interface designs for mobile devices.

Given the lack of prior multilingual research on mobile search interfaces, the objective of this phase is to develop and evaluate an initial set of mobile multilingual news search interface designs. The three research questions for this phase are: (RQ3.1) What type(s) of
mobile multilingual news search interface designs are preferred by polyglot users? (RQ3.2) What are the interaction behaviors of polyglot users with mobile multilingual news search systems? (RQ3.3) Do different mobile multilingual search interfaces and/or news topics lead to different user behaviors?

1.2.4 Phase 4: Comparing user preferences and behaviors between desktop and mobile multilingual news search interfaces.

The purpose of this phase is to evaluate different interactive multilingual news search interface designs to better understand how best to support multilingual web users and their multilingual skills, both on desktop and mobile. The three research questions for this phase are: (RQ4.1) What type(s) of interactive multilingual news search interface designs are preferred by polyglot users? Specifically, what type(s) of interfaces are preferred on desktop and mobile devices? (RQ4.2) What are the interaction behaviors of polyglot users with different interactive multilingual news search systems, both on desktop and mobile devices? (RQ4.3) Do different interactive multilingual search interfaces and/or news topics lead to different user behaviors?

1.3 Methodology

To answer the above research questions, this research provides the first in-depth user-centered investigation of multilingual news search interfaces through four study phases, which include surveys, interviews, as well as task-based user studies using both crowdsourcing and laboratory experiments. Across all 4 phases, a total of 2875 participants took part in these studies.

1.3.1 Phase 1: Multilingual Search Engine User Interface Study & Pilot Multilingual News Consumption Behavior Interview

This phase included two user studies, which provide the motivation to study multilingual news search.

1.3.1.1 Phase 1.1: Multilingual Search Engine User Interface Study - Laboratory Experiment with Eye Tracking

The first part was a comparative user study of interactive multilingual search interfaces. Specifically, this experimental study provided the first in-depth comparative analyses of user behaviors and preferences with respect to four different multilingual search user interfaces, through a lab-based experiment involving 25 participants.
1.3.1.2  Phase 1.2: Pilot Multilingual News Consumption Behavior Interview

The second part was a small-scale interview study, which investigated current multilingual user news browsing/selection behaviors. Specifically, this study aimed to gain an initial understanding of current multilingual user news checking, browsing and reading behaviors through a semi-structured interview with 6 participants.

1.3.2  Phase 2: Survey and Task-Based User Study (Through Crowdsourcing)

This phase included two large-scale user-centered studies, aiming to obtain a better understanding of multilingual news consumption behaviors, i.e. how multilingual users search/browse/check news in multiple languages.

1.3.2.1  Phase 2.1: Multilingual News Consumption Behavior Survey

The first study was a survey that was distributed through the FigureEight\(^6\) crowdsourcing platform (originally known as CrowdFlower). This survey consisted of a questionnaire that enquired about polyglot users’ current news consumption behaviors. Specifically, the questionnaire included questions related to platform usage frequency, access modality, news type, as well as shortcomings of current news access systems. In total, 1200 participants took part in this survey.

1.3.2.2  Phase 2.2: Task-Based User Study

This user study specifically investigated user query language choices and result language choices with respect to news search. Specifically, four separate tasks elicited query formulation (Task 1), query selection (Task 2), and search result choices (Task 3, Task 4). For each of these tasks, the FigureEight crowdsourcing platform was used to gather a pool of 1600 participants.

1.3.3  Phase 3: Lab-Controlled Interview

As previously mentioned, the goal of phase 3 was to research initial designs of mobile multilingual news interfaces. This phase therefore first consisted of the design and development of an initial set of mobile news search interface prototypes, followed by a pilot user study to gauge initial user reactions. This phase included a total of 10 participants.

\(^6\)Figure Eight: https://www.figure-eight.com
1.3.4 Phase 4: Task-Based User Study (Through Laboratory Experiment)

Phase 4 consisted of the refinement of the mobile prototypes first trialed in phase 3, followed by a comparative lab-based experiment that evaluated both desktop and mobile multilingual news search interfaces. This phase included a total of 40 participants.

1.4 Contribution

The first scientific contribution of this thesis is the establishment of design guidelines for multilingual information/news access systems for desktop and mobile devices, based on the understanding of how to best support polyglot users and their multiple language skills.

The second contribution consists of the first large-scale investigation of current multilingual news consumption, as well as a detailed analysis of multilingual querying and search result selection behaviors.

These contributions have led to high-quality conference and journal publications:


- (In progress) Ling, C., Steichen, B., & Figueira, S. A Comparative User Study of Interactive Multilingual News Search Interfaces on Desktop and Mobile.

1.5 Thesis Overview

Chapter 1 introduces the motivation, objectives, and methodology for this dissertation work, and in addition it provides a brief overview of the later chapters. Chapter 2 summarizes the related research work of multilingual news search and interfaces. Chapter 3 presents a comparative user study of interactive multilingual search interfaces (phase 1 study). Chapter 4 presents an investigation of consumption, querying, and search result selection behaviors (phase 2 studies). Chapter 5 presents the design and evaluation of different multilingual news search interface designs for mobile devices (phase 3 study). Chapter 6 compares user
preferences and behaviors between desktop and mobile multilingual news search interfaces (phase 4 study). Chapter 7 describes the conclusion and future work.
Chapter 2

Related Work

While prior work has not directly studied multilingual user behaviors with respect to online news access, there are a number of related research studies that have investigated similar issues related to multilingualism. In particular, the areas that are most directly related to the work in this thesis are general Multilingual Search, as well as Multilingual News summarization.

2.1 Multilingual Search

The majority of research in general multilingual search has concentrated on Cross-language Information Retrieval (CLIR) and Multilingual Information Access (MLIA), which mainly focus on improving translation accuracy and retrieval effectiveness (e.g. as part of evaluation forums such as CLEF\(^7\)) [4,5].

In terms of interactive cross-language information retrieval, research has generally found that users make an increased use of translation tools for low-proficiency languages, while for familiar languages, users tend to write translated queries by themselves [6,7]. Moreover, it has been suggested that CLIR systems should provide “simple” and “intuitive” translation interaction functions, for example, auto-translation tools that allow users to edit and keep track of prior queries [6,8].

A number of survey-, interview-, and focus group-based studies have also been conducted to investigate general multilingual information seeking and browsing [4,7,9-12]. For example, the work in [10] explored multilingual user behaviors in a digital library context and found that users typically make use of different languages depending on their language skills and field of knowledge. Several qualitative studies [9,12] also found similar results, namely that the main challenges faced by multilingual users relate to language proficiency issues in terms of query formulation and reformulation in secondary languages (e.g., if a user’s second language writing ability is limited). Likewise, a large survey in [11] studied general usage behaviors of multilingual users when browsing and searching the Web. This study similarly found that multilingual users generally tend to use many of their languages online, but also

\(^7\)CLEF (Conference and Labs of the Evaluation Forum): www.clef-initiative.eu/
that proficiency and type of content have an influence on their selections. Several studies [13-15] have also researched language/code-switching behaviors in online searching, where code-switching is typically defined as “switching of languages during conversation” [16-19]. For instance, the research in [13,14] studied reasons for situational and metaphorical code-switching episodes. Users were found to use situational code-switching when search tasks required them to change languages to obtain efficient and adequate search outcomes, and it appeared to be influenced by pragmatic reasons linked to information sufficiency, language proficiency, information verification, context, and translation. On the other hand, metaphorical code-switching behaviors occurred when users tended to get satisfactory search outcomes, owing to the reasons of a language’s perceived image (e.g., accurate and objective), cultural perspective, website credibility, sense of belonging, and psychological acceptance. Findings in [15] suggested that the usage of mixed language queries relied not only on the topic and information availability, but it was also influenced by how users expected the topic to be represented in different cultures. The author also emphasized that mixed language queries are irreplaceable to serve certain information needs, where single language queries or other search features cannot match the quality of the outcome obtained from mixed language queries. However, the research found that current search engine systems do not support these multilingual user search behaviors well. For example, Google’s language search setting feature was found to not replace mixed language queries for three reasons, including inconvenient interface design (difficult to discover), unfamiliarity of the feature, and ineffectiveness of retrieving cross-language resources. Most similar to our work, the research in [4] consisted of a large-scale user study that explicitly explored general web search behaviors of multilingual users, showing that users tend to make significant use of each of their languages, particularly when the system provides additional support, such as auto-translation of queries.

In terms of the design of systems that present aggregated content, most research has so far been conducted on monolingual aggregated search interfaces that display multiple content types/verticals (e.g. image, news, video) [5,20,21]. Only a limited number of human-centered studies have been conducted to better understand and support the abilities and preferences of multilingual users. Most notably, the studies in [8,22] have investigated different multilingual search result pages for general web search (e.g. interleaved multilingual result lists versus separated language panels). This work has uncovered several interesting initial results, particularly the finding that multilingual users seem to prefer strict language separation (e.g.}
as provided by separate panels), and that the traditional approach of interleaving results, as typically used in prior research as well as current popular search engines (e.g., Google, Bing, Baidu) is least preferred. Lastly, Chu et al. [8] conducted a qualitative user study of a search engine which allowed users to customize multilingual result layouts from different search engine platforms (e.g., English results from Bing.com, Simplified Chinese from Baidu.com). Again, aggregated layouts were shown to help users to compare retrieval results easily, and they noted that “simplicity”, “visibility”, and “customization” were considered the most important characteristics of a multilingual web search system. However, none of these studies focused on current news consumption behaviors by multilingual users, nor querying and result selection behaviors with respect to multilingual news content. In addition, no research has focused on such systems on mobile platforms.

2.2 Multilingual News

In terms of prior work in the area of multilingual news, the majority of projects have focused on multilingual news summarization and clustering, or the development of monitoring platforms [23-29].

For example, the Columbia Newsblaster news summarization system developed by Evans et al. [24] aggregates news from multiple languages across multiple sites. Specifically, the system automatically collects and summarizes news from multiple source languages, enabling users to browse news topics through English summaries, and compare different viewpoints from different countries. Similarly, Chung et al. [23] developed a system called the Chinese Business Intelligence Portal (CBizPort), which is an integrated news search engine platform allows user specifically simultaneously searches, summarizes, and categorizes multilingual (Simplified Chinese and Traditional Chinese) business news information from different regions in China, such as mainland China, Taiwan, and Hong Kong. Wan et al. [25] investigated the task of multilingual news summarization to find and summarize significant differences between news articles about the same event in English and Chinese, while Chen et al. [26] proposed a general-purpose architecture of a multilingual news summarizer, which includes monolingual and multilingual clustering, as well as the presentation of summarization results.

Also related to the topic of multilingual news access, the JRC’s Europe Media Monitor (EMM) team built a real-time multilingual news monitoring and analysis system [28,29], which clusters top events from multilingual sources in the European Union (EU) and beyond.
The EMM conducted large-scale media analysis in up to 70 languages, recognizing various types of trends, and some of which were generated using combined information from news articles written in different languages and from different social media portals. It also allows users to explore multilingual information through interactive maps and graphs, enabling users to evaluate the data from different viewpoints [27].

Despite the above efforts in terms of building multilingual summarization or monitoring systems, there has been a distinct lack of human-centered research in the area of multilingual news. In particular, prior research has so far not investigated multilingual user information needs, preferences, and behaviors, nor the relationship of multilingual news access for polyglots of varying proficiency levels and across different topic domains. Likewise, the majority of systems so far have focused on summarizing news in a single language, rather than aggregating the original content, which would allow users with multiple language skills to access original and unaltered multilingual information.
Chapter 3

Phase 1: Multilingual Search Engine Preferences and Behaviors & News Consumption Behavior Pilot

3.1 Phase 1.1 A Comparative User Study of Interactive Multilingual Search Interfaces

As outlined in the previous chapter, one of the greatest unexplored challenges lies in how to best support polyglot users from an interface and interaction perspective. While many general search systems technically already support multiple languages, they typically tend to emphasize distinctions between languages, often requiring users to switch between systems, or at least conduct separate searches to obtain results from more than one language source. This extra effort reduces the probability of using multiple languages, which may result in less relevant content being found (especially given the language imbalance discussed above). To conquer these shortcomings, there has been significant initial progress in building systems that can retrieve information from multiple languages. In particular, cross-language search systems have been developed to allow users to search for information in a language that is different from the query language. However, there has been a distinct lack of research on the user interaction aspects of general multilingual search, especially in terms of how to design interactive multilingual search result interfaces.

Before focusing specifically on multilingual news browsing and search, the objective of phase 1.1. is therefore to evaluate different general multilingual search interface designs, to get a better understanding of how to best support polyglot Web users and their multiple language skills. Specifically, the following research questions are asked:

- (RQ1.1.1) What type(s) of interactive multilingual search interface designs are preferred by polyglot users?

- (RQ1.1.2) What are the interaction behaviors of polyglot users with interactive multilingual search systems?

- (RQ1.1.3) Do different interactive multilingual search interfaces and/or task types lead to different user behaviors?
In order to answer these research questions, the study in phase 1.1. provides the first in-depth comparative analyses of user behaviors and preferences with respect to four different multilingual search user interfaces, through a lab-based user study involving 25 participants.

3.1.1 Experimental Setup

This section first describes in detail the four different interfaces that were designed and developed for our study. These interfaces were based on the techniques identified from prior work, and in particular they extend the static multilingual interfaces presented in [22]. This is followed by a description of the experimental platform, study tasks, and user study and analysis procedure.

3.1.1.1 Interfaces Used in the Study

**Tabbed Interface**

As shown in Figure 1, the ‘Tabbed’ interface allows users to change between lists of results in different languages through tabs. The language tabs reflect the languages that the user is proficient in (up to 4 languages), as determined through the pre-questionnaire. As with each of the other interfaces described below, the ‘Tabbed’ interface automatically translates (using the Microsoft Translator API) the user’s query to the language of the currently selected tab. For example, in Figure 1, a user with English and Chinese proficiency has typed an English query “recipes for chocolate pudding”, and has then chosen the Chinese Tab to view Chinese results, which are retrieved using the auto-translated query “巧克力布丁的食谱”. Translations can also be edited, as recommended in prior work [30], and several translation alternatives are provided as suggestions to the user. Each result page displays a single monolingual ranked list, which contains 12 results (retrieved using the Bing Search API). At the bottom of the page, participants can click ‘previous’ and ‘next’ buttons to view more search results.

**Panels Interface**

The ‘Panels’ interface (Figure 2) displays one or more panels on a single result page, with each panel displaying results in a different language. The number of languages (and hence panels) per page is determined by the user through the use of checkboxes (up to 4 languages).

---

8datamarket.azure.com/dataset/bing/microsofttranslator - While there are many different machine translation techniques and tools, this was not the focus of this study. The Microsoft Translator API supports a large number of language pairs, and it was found to work sufficiently well for the topics used in the study tasks.

9datamarket.azure.com/dataset/bing/search
A user may choose to view results in each of his/her languages, or only focus on one single language.

To make sure that the same number of results is displayed for each result page, and to ensure consistency across interfaces (e.g. ‘Panels’ having the same number of results per page as ‘Tabbed’), the number of results per language equals to 12 divided by the number of languages. For example, if a user has selected 4 languages, each individual panel will contain 3 results. Conversely, if only one language is chosen, the result list is equivalent to the ‘Tabbed’ interface, i.e. 12 results for the chosen language. Each panel is headed by the (editable) auto-translated query according to the respective language, and the results within each panel can be navigated with ‘previous’ and ‘next’ buttons.

![Figure 1 Tabbed Interface](image1)

![Figure 2 Panels Interface](image2)

**Interleaved Interface**

As shown in Figure 3, the ‘Interleaved’ interface has a single result list, with Web search results from different languages being interleaved. Checkboxes again allow users to select/deselect results in specific languages, and all (editable) auto-translated queries are
shown together above the interleaved list. The number of results per page is again 12. For example, if a user has selected 2 languages (as in Figure 3), there will be 6 results per language, with result 1 being in the user’s first language, result 2 being in the user’s second language, result 3 again in the first language, etc.

![Figure 3 Interleaved Interface](image)

**Universal Interface**

The ‘Universal’ search interface also presents a single result list similar to the ‘Interleaved’ interface. However, results from each language are grouped together into blocks, rather than interleaved with each other. For example, as shown in Figure 4, when a user selects 2 languages, a block of 6 results in the user’s first language are shown above a block of 6 results in the user’s second language.

![Figure 4 Universal Interface](image)
3.1.1.2 Platform and Participants

A lab-based study was conducted using a purpose-built experimental platform (built using PHP and MySQL), running on a desktop computer with a 24-inch monitor. In addition, user eye gaze was tracked using a Tobii X3-120 eye tracker, which is a remote eye tracker providing unobtrusive tracking.

Experiment participants were recruited through campus mailing lists, and were required to have ‘some proficiency’ in at least two out of a set of 7 languages (English, French, German, Spanish, Italian, Simplified Chinese, Traditional Chinese). Each participant was given a gift voucher worth 20 USD, and the average time to complete the entire study was approximately 90 minutes.

3.1.1.3 Task Questions

Each user was presented with a total of 12 tasks (P-value<0.), which were to be completed using the 4 interfaces (3 tasks per interface, counterbalanced across participants/interfaces). These tasks were sourced from CLEF (2006-2011) and TREC (2012-2014) campaigns, representing a variety of general Web search topics. The tasks were manually categorized into the following three different task type categories (using the definitions/categorization schemes in [31]: Doing, Fact Finding, and Learning tasks, and all topic descriptions were manually translated to the seven supported languages prior to the study using native speakers.

Before performing assigned tasks with each interface, participants first performed a practice task (the same practice task was used for each interface).

3.1.1.4 Procedure

There were five main steps each participant followed in the experiment. These five steps were i) calibrating the eye tracker ii) completing an initial questionnaire, iii) performing tasks, iv) completing mid-study questionnaires, and v) completing a post-study questionnaire.

The first part was a demographic questionnaire that included self-reported language proficiency in terms of reading, listening, and writing ability. Proficiency was measured on a scale from 1 to 4, with 1 corresponding to very limited proficiency, and 4 corresponding to native proficiency.

10Having ‘some proficiency’ was defined as ‘you must have some reading/writing ability’ in the language.
Table 1 Task Questions and Question Types

<table>
<thead>
<tr>
<th>Question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find recipes for making chocolate puddings.</td>
<td>Doing</td>
</tr>
<tr>
<td>You want to know how a .csv file can be imported in excel.</td>
<td>Doing</td>
</tr>
<tr>
<td>You want to buy Yves Saint Laurent boots. You want to find places to buy, reviews, etc.</td>
<td>Doing</td>
</tr>
<tr>
<td>Rock climbing for beginners only publications which specifically provide information on climbs that are not difficult or give instructions on rock climbing for beginners are of interest.</td>
<td>Doing</td>
</tr>
<tr>
<td>What is the current price of oil?</td>
<td>FactFinding</td>
</tr>
<tr>
<td>Give the names and/or location of places that have been designated as UNESCO World Heritage Sites of outstanding beauty or importance.</td>
<td>FactFinding</td>
</tr>
<tr>
<td>What conditions trigger asthma in children?</td>
<td>FactFinding</td>
</tr>
<tr>
<td>How high above ground level is the ozone layer?</td>
<td>FactFinding</td>
</tr>
<tr>
<td>Look for information on the existence and/or the discovery of remains of the seven wonders of the ancient world.</td>
<td>Learning</td>
</tr>
<tr>
<td>Find publications providing general introductions to food allergies and the prevention of such allergies.</td>
<td>Learning</td>
</tr>
<tr>
<td>We seek any information on human cloning including claims of the production of the first human clone.</td>
<td>Learning</td>
</tr>
<tr>
<td>In what sports are drugs used illegally?</td>
<td>Learning</td>
</tr>
</tbody>
</table>

Next, users interacted with the previously described interfaces. More specifically, for each interface, users performed one practice task, followed by 3 tasks. For each task, participants were asked to issue their own queries to look for information on the given topic, with the topic being displayed in all of the participant’s languages. Participants were also asked to mark any results they found relevant or helpful in learning about or solving the task, using buttons placed next to each result (see Fig. 1-4). These markings were recorded by the experimental platform, along with task times, queries issued, as well as documents viewed (i.e. results clicked/opened). Participants were free to move on to the next task if they felt that they had received sufficient relevant information on the task (in line with similar studies in aggregated search [20,21,32].

Upon completion of all 3 tasks with one interface, a mid-study questionnaire was presented, which asked users to indicate their perceived level of search support provided by the interface.

Lastly, after having interacted with all 4 interfaces, a post-study questionnaire asked participants to indicate their overall comparative impressions, including i) which interface was ‘easiest to use’, ii) which interface they ‘preferred the most’, and iii) which interface they
‘disliked the most’. In addition, participants were asked to indicate their reasoning for particular preferences/dislikes.

3.1.1.5 Data Analysis

To understand multilingual user behaviors and preferences, both user interaction data and user preference data were analyzed. For each of these analyses, statistical significance will be reported as determined through Chi-squared tests or ANOVA tests (as well as Bonferroni-corrected posthoc tests).

In addition, several types of data captured through eye tracking were analyzed, including number of fixations, fixation rate, fixation duration, and saccade length (see Table 2 for feature descriptions).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation number</td>
<td>Total number of eye gaze fixations</td>
</tr>
<tr>
<td>Fixation time</td>
<td>Total fixation time spent</td>
</tr>
<tr>
<td>Fixation rate</td>
<td>Fixation number divided by time interval (e.g., fixations per second)</td>
</tr>
<tr>
<td>Saccade length</td>
<td>Distance between fixations</td>
</tr>
</tbody>
</table>

3.1.2 Results

This section first presents participant demographics, followed by the results for user preference ratings. Next, results for interaction behaviors (e.g. task time, number of queries, etc.) will be presented. Lastly, a detailed analysis of differences in interaction behaviors between the different interfaces is presented, including analyses of eye gaze data for each interface, as well as influences of different task types.

3.1.2.1 Participant Demographics

In total, 36 participants took part in the study, of which responses were retained from 25 after filtering out cases of recording errors.

The 25 retained participants had an age range of 19-55, and, in accordance with the study requirements, were all proficient in at least two languages. 18 participants indicated that they were proficient in two languages, 5 participants were proficient in three languages, and 2 participants in four languages. The most common first languages were English (16), Chinese (8), and German (1). The most common second language was Spanish (10), followed by English (9), French (4), and Chinese (2). All participants indicated native proficiency in their
first language, and the majority of participants indicated high proficiencies (level 3 or 4) in their second language across reading, writing, and listening (22/18/20 respectively).

3.1.2.2 User Preferences (RQ1.1.1)

As previously mentioned, user preferences were gathered through a comparative post-study questionnaire (i.e. after all tasks with all interfaces had been completed), as well as mid-study questionnaires (after all tasks for an individual interface were completed).

![Post-study questionnaire results. ‘Panels’ is considered the ‘most preferred’ and ‘easiest to use’ by most participants, while ‘Interleaved’ is the ‘most disliked’.

In the post-study questionnaire, participants generally indicated clear preferences/dislikes towards specific interfaces, with the interface type having a statistically significant effect on all post-study questionnaire answers. As shown in Figure 5, the ‘Panels’ interface was strongly considered to be the ‘easiest to use’ compared to all other interfaces (p-value<0.001), and the comparatively ‘most preferred’ (p-value<0.001) interface. The ‘most disliked’ interface was the ‘Interleaved’ interface (p-value<0.01).

For the mid-study questionnaires (i.e. after having finished all tasks with one particular interface), results similarly indicated differences between interfaces, although less pronounced, as shown in Figure 6. In particular, differences between interfaces were found to be statistically significant for Q5 (‘getting an overview’), and marginally significant for Q3 (‘helping combine information’). Moreover, through Bonferroni-corrected post-hoc tests, it was found that the ‘Panels’ interface was considered to be particularly good at ‘getting an
overview’ compared to the ‘Tabbed’ and ‘Universal’ interfaces (p-value<0.05 and p-value<0.09 respectively).

Q1: “The system provided enough information to help me solve the search tasks.” (p-value<0.1)

Q2: “The system provided me with many different kinds of information.” (p-value<0.4)

Q3: “The presentation of search results helped me easily combine information from multiple languages.” (p-value<0.07)

Q4: “The presentation of search results allowed me to easily identify relevant information.” (p-value<0.1)

Q5: “The presentation of search results helped me get an overview of the information available in multiple languages.” (p-value<0.02)

Figure 6  Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree).

Additional analyses with respect to participants’ proficiency levels did not yield any statistically significant results, suggesting that the preferences were not influenced by participants’ language skills in their secondary languages. However, as previously mentioned, language proficiencies of participants were not well distributed (the vast majority of participants indicating a proficiency level of 3 or higher in their secondary language), hence requiring further research with a participant base that includes lower second language proficiencies.

3.1.2.3 General User Behavior (RQ1.1.2)

Query behavior

User queries were analyzed in terms of two criteria, namely i) number of queries, and ii) languages used for queries.
There were 557 queries in total, most of which were entered in the main search box. Surprisingly, only 3 users used the edit query function to revise auto-translations, editing a total of only 4 queries. On average, participants issued 1.86 queries per task, and 22.28 queries for the whole study session.

In order to analyze users’ language choices, queries were divided into 3 types, depending whether the query contained a user’s first language (L1), a secondary language (L2), or both L1 and L2 (Mix). There was a statistically significant difference (Figure 7) in terms of language use (p-value<0.001), with participants performing L1 queries more often than L2 queries, and rarely using mixed-language queries. Nevertheless, as shown in Figure 7, it is noteworthy that the number of L2 queries was still high (approximately half as many as L1 queries). However, there were 6 out of 25 participants who only used L1 during their whole study session.

![Figure 7](image)

**Figure 7** Average query number per language (Mix indicates the use of multiple languages within a single query)

When analyzing query language changes within all 300 tasks (25 users * 4 interfaces * 3 task types), it was found that (as Table 3) during a total of 240 tasks (80%), participants only used a single language for querying, i.e. only L1, only L2, or only Mix (with L1, a user’s first language, being the most popular). Conversely, during 60 tasks (20%), participants switched between languages for querying (e.g. starting with L1 queries, then switching to L2).
Table 3  Query language changes within tasks

<table>
<thead>
<tr>
<th>Query Pattern</th>
<th>Language Sequence</th>
<th>Percentage</th>
<th>Percentage Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Change</td>
<td>L1 only</td>
<td>62%</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>L2 only</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mix only</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>L1 -&gt; L2</td>
<td>10.67%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L2 -&gt; L1</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1 -&gt; Mix</td>
<td>1%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Mix -&gt; L2</td>
<td>0.33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple switches</td>
<td>5%</td>
<td></td>
</tr>
</tbody>
</table>

Result viewing and marking behavior
In addition to queries, the number of documents viewed by participants (i.e. result links that participants had clicked on) was also analyzed, as well as the number of documents that participants had marked as relevant to the task.

On average, participants clicked on 7.98 documents per task. When breaking down by language, the number of documents viewed in participants’ L1 was slightly higher at 4.54 compared to L2 at 3.42. However, this difference was not found to be statistically significant. Compared to the query results above, participants made more use of each of their languages during tasks, with participants choosing to click on results in each of their languages during 44% of the tasks. Conversely, only during 38% of tasks did participants click on results in only their L1.

The average number of documents marked as relevant was 6.25, and, similar to the above results on documents viewed, there was only a small difference between L1 and L2 (average of 3.43 vs 2.82). Again, this difference was not statistically significant. Participants very often marked results as relevant in each of their languages (74% of tasks). Conversely, only during 20% of tasks did participants mark results in only their L1 as relevant.

Use of language checking/unchecking and tabbing
As discussed in section 3.2, three interfaces (‘Panels’, ‘Interleaved’, ‘Universal’) allowed users to check/uncheck languages during their searches (e.g. to only search/display results in L1). Participants only made use of this functionality during practice tasks (where they were explicitly explained the functionality and instructed to try it), and not at all during the actual tasks. The number of tab switches for ‘Tabbed’ was also low, with only an average of one tab switch per user per task.
3.1.2.4 Influence of Interface and Task types (RQ1.1.3)

Comparison of Interface Types

Overall, participant behavior differences in terms of average task time, number of queries, number of documents viewed, and number of documents marked as relevant were not found to be statistically significant (p-value<0.08/p-value<0.31/p-value<0.14 respectively). Likewise, no difference was found in terms of query lengths. However, there appeared to be a common trend (and near statistical significance) towards longer task times, higher number of queries, and higher number of documents marked as relevant for the ‘Tabbed’ interface compared to all other interfaces (see Figure 8 for task time). While ‘Interleaved’, the most disliked interface, showed the shortest task time, this was not found to be statistically significant.

![Figure 8: Average task time per interface](image)

Three eye-tracking variables, namely fixation number, fixation time, and saccade length, did show statistically significant differences with respect to interface (p-value<0.0085/ p-value<0.0003/p-value<0.002 respectively), confirming the trends observed in the above behavior differences. In particular, the distributions (Figure 9) were similar to task time, where ‘Tabbed’ had the highest value.

Through post-hoc tests, it was found that participants had a much higher fixation number on the ‘Tabbed’ interface compared to the ‘Universal’ interface (p-value<0.01). For fixation time, ‘Tabbed’ had statistically significantly longer fixation time than ‘Interleaved’ (p-value<0.05) and ‘Universal’ (p-value<0.008). Regarding saccade lengths, however, the ‘Tabbed’ interface had shorter length compared to ‘Interleaved’, ‘Panels’ and ‘Universal’ (p-value<0.06/p-value<0.005/p-value<0.005). Additionally, ‘Panels’ also had longer saccade
length compared to ‘Interleaved’ (p-value<0.04). Fixation rate did not yield any statistically significant differences for interface type (p-value<0.05), but ‘Tabbed’ still appeared to have the highest values. Again, these measures confirm the aforementioned trends that the ‘Tabbed’ interface required users to spend more time.

In addition to statistically comparing the raw eye gaze measures, we also generated visual heat maps for the four interfaces\textsuperscript{11}. More specifically, we generated ‘first glance’ heat maps, which visualized participants’ initial fixations before performing their first result view click\textsuperscript{12}.

![Average Fixation Number](image1)
![Average Fixation Time](image2)
![Average Fixation Rate](image3)
![Average Saccade Length](image4)

**Figure 9  Eye tracking features per interface**

The heat maps shown in Figure 10 are constructed through gradient colors from green to red, which represent eye fixation counts from low to high. The heat maps generally follow the same pattern as found in monolingual search research, showing a strong skewing towards the top few results. This meant that for ‘Universal’, L2 results received relatively less attention (although not resulting in statistically significantly lower document views). The heat

\textsuperscript{11}These maps were generated from the number of eye gaze fixations from all the participants who had indicated two languages. Heat maps across participants with different numbers of languages cannot be overlaid given the slightly different overall page lengths and layouts. The number of participants with more than two languages was very low, hence these participants were not analyzed separately.

\textsuperscript{12}Overall heat maps from users’ entire interactions (i.e. beyond ‘first glance’) yielded very similar results.
map for the ‘Panels’ Interface reveals that participants frequently checked the top results for both ‘L1’ and ‘L2’, suggesting that both panels were regarded as equally important. This pattern also explains the higher saccade lengths found above.

**Figure 10  First glance heat maps**

*Influence of Task Type on User Behavior*

Lastly, a series of additional analyses was conducted to investigate the influence of task type (*Doing/Fact Finding/Learning*), as well as any interaction effects between task type and interface type. A statistically significant effect was found for task type on task time (p-value<0.001), with *Learning* tasks generally taking longer than both *Fact Finding* and *Doing* tasks.

In addition, there was a statistically significant difference regarding the number of L1 queries with respect to task type (p-value<0.003), with *Learning* tasks having a higher number of L1 queries than *Doing* and *FactFinding* tasks. Interestingly, there was no statistically significant difference in the number of L2 queries (p-value<0.35) or Mix queries (p-value<0.31) with respect to task type. Figure 11 illustrates this influence of task type on the use of L1, L2, and Mix queries. As can be seen in this graph, participants used a higher number of L1 queries during *Learning* tasks (particularly compared to *Doing* tasks), whereas the number of L2 queries was more stable.
Related to this, it was found that participants changed the query language less during FactFinding tasks compared to both Doing tasks (p-value<0.003) and Learning tasks (p-value<0.01). Similar results were found for the number of documents marked as relevant. However, no statistically significant differences could be found.

Lastly, there were no interaction effects between task type and interface type for any of the measures.

### 3.1.3 Summary and Discussion

This section summarizes the key findings from our study regarding interface-specific user preferences and interaction behaviors (RQ1.1.1 and RQ1.1.3), as well as general interaction behaviors found across all interfaces (RQ1.1.2). In addition, the section will compare these results to prior work, provide further insights and recommendations based on participants’ free text responses (given as part of the post-study questionnaire), and discuss implications for designing multilingual search interfaces.

The ‘Panels’ interface was by far the most preferred interface, and it was regarded as providing the best ‘overview of information’. Additionally, when analyzing participants’ free text responses, it was found that the ‘Panels’ interface was considered ‘very organized’ due to the ‘clearly separated languages’. Participants reported that this made it easier to differentiate between the different languages, allowing one to focus on one language at a time. In addition, compared to the ‘Tabbed’ interface, which actually had the clearest language separation, participants liked that the ‘Panels’ interface still allowed the viewing of results in each language without any additional interface actions (e.g. clicking between tabs). This strong...
preference for the ‘Panels’ interface is much more pronounced than in related monolingual aggregate search [20,21,31], hence suggesting that the combination of results in multiple languages is different from the combination of different verticals in traditional aggregated search (e.g. Web and image results). This may stem from strong saliency differences in traditional monolingual aggregate search (e.g. image results having a strong visual saliency compared to textual Web search results), which different multilingual results are lacking. In addition, this preference for ‘Panels’ is even more pronounced than in our prior work that used only static multilingual interfaces [22]. However, several users still reported some issues interacting with the ‘Panels’ interface. In particular, some users did not feel comfortable having to ‘move their eyes around too much’ to view different language panels when reading. In terms of behavioral differences, this increased gaze effort is supported by the saccade lengths found in the eye tracking analysis, which showed that the panels interface does induce longer saccades (although not increasing task times).

The ‘Interleaved’ interface, which interleaves results from different languages in a single ranked list, was found to be the ‘most disliked’ interface. In participants’ free text comments, it was noted that the proximity of the different languages made it difficult for participants to ‘transfer their reading mode’ and thus they ‘could not focus’ and ‘felt confused’. This may again be due to the aforementioned lack of saliency differences between different multilingual results, leading to extra cognitive effort for participants to ‘code-switch’ between languages. However, despite its general unpopularity, there were still some users who preferred the ‘Interleaved’ interface, stating that they thought it was ‘faster to read’. This individual user differences may hence partially explain why there were no major behavioral differences (e.g. for task time) between ‘Interleaved’ and other interfaces.

While interfaces with clear language separation were generally preferred, the ‘Universal’ interface, which separates languages in vertical blocks, had the lowest vote for the most preferred, and the lowest number for the most disliked (hence evoking no strong reaction either way). Several participants noted that L2 results were too far down the list, meaning that they were sometimes not noticed. Nonetheless, the interface did not suffer from the strong general dislike that the ‘Interleaved’ interface received, which is different from prior findings with static interfaces [22]. It therefore appears that, if used in an interactive setting, ‘Universal’ is generally the preferred ‘vertical result list’ interface.
The most controversial interface was ‘Tabbed’, as it was voted second for ‘easiest’ and ‘preferred’ interface, yet also second for ‘most disliked’. It appears therefore that there are clear differences between users. Some participants preferred ‘Tabbed’ to all other interfaces because of the ability to fully switch between languages (and hence having more results in the desired language on one page). However, others disliked it compared to all other interfaces because they had to switch manually between languages, which made it difficult to get an overview. In terms of behavioral differences, there were several trends suggesting that a ‘Tabbed’ interface leads to longer task times, an increased number of queries, an increased number of documents marked as relevant, as well as a higher number of fixations and fixation durations. These trends echo similar results found for monolingual aggregate search [21], where it was found that ‘Tabbed’ interfaces generally afforded looking deeper into the details of individual result sources, while other ‘Universal’ (called ‘Blended’) interfaces afforded getting a quick overview of multiple sources.

In terms of general user interaction behaviors, there were several interesting findings across all interfaces. First of all, most participants made significant use of each of their languages during the study session, reinforcing the general motivation for providing systems that explicitly support multilingual users. While participants made less use of their L2 during querying, they did view and mark L1 and L2 documents almost to an equal degree. This suggests that when users choose to enter queries in a particular language (e.g. due to higher proficiency or personal preference), they are often still interested in viewing results in their other languages. This again confirms the usefulness of search result pages that integrate results from multiple languages. Interestingly, participants in our study did not make much use of query editing capabilities as recommended in prior work [9], nor language selection/deselection. This finding suggests that users prefer to simply reissue a new overall query, rather than interacting with a subpart of the multilingual result page. Participants’ eye gaze behavior for each of the interfaces was generally similar to monolingual search result page results, with top results receiving comparably higher numbers of fixations. A notable pattern was shown for the ‘Panels’ interface, where participants gave equal priority to each individual panel. This suggests that horizontal panel placement may be of great importance when presenting multilingual result pages.

While the effect of task type was also analyzed, not many statistically significant differences were found. Most notably, it was found that for Learning tasks, users issued an increased number of queries, and this increased number appeared to be mostly additional L1
queries. Conversely, for Doing and Fact-Finding tasks, which require less effort in general, the distribution between L1 and L2 was more balanced. This suggests that when users go beyond an initial overview of results (often sufficient for Doing and particularly simple Fact-Finding tasks), they tend to stick with their L1. Therefore, it may be sufficient to display only L1 results beyond the first result page. This would be similar to aggregated search practices in modern search engines, which often do not display aggregated pages beyond the first result page.

Lastly, across all interfaces, many participants noted in free text comments that it was important to clearly mark the language of individual or groups of results, e.g. through explicit labels. This need for clarity may have contributed to the general dislike of the ‘Interleaved’ interface, as it is lacking clear signifiers as to the language of each result. In addition, some participants indicated that they sometimes did not notice the different language sections, or completely forgot that other languages were available. This was particularly noted for the ‘Universal’ interface, suggesting that result blocks in secondary languages should potentially be ‘slotted’ higher in the vertical search result list in order to get noticed. This may take the form of modern aggregated search interfaces, where different verticals are slotted into a ‘main list’ dynamically. This variation of the ‘Universal’ interface warrants further research, as it may help users better notice the different languages, while still maintaining a clear language separation, which was generally found to be beneficial.

3.1.4 Conclusion

While the population of multilingual Web users across the globe has increased dramatically, there has been a distinct lack of research on how to best support multiple language abilities. To address this gap, this section has provided the first evaluation of different interactive multilingual search interfaces. In particular, these interfaces were designed specifically with multilingual users in mind.

Overall, it was found that clear language separation was strongly preferred, whereas an ‘Interleaved’ approach was the least preferred. The findings from this interactive study thereby extend similar results found in studies that used static interfaces, and further present a compelling case for changing how multilingual search interfaces are built. In particular, while prior research in Multilingual Information Access has traditionally been confined to ‘Interleaved’ or ‘Tabbed’ presentation approaches, the study has shown that these approaches are in fact often the least preferred by multilingual users. In terms of interaction behaviors, it
was found that users made significant use of each of their languages during search sessions. In particular, participants viewed and marked documents as relevant to a similar degree in each of their languages.

3.2 Phase 1.2 Pilot Multilingual News Consumption Behavior Interview

3.2.1 Introduction and Research Questions

While phase 1.1. had investigated general Web search interfaces to support multilingual users, phase 1.2. was conducted to gain an initial understanding of multilingual users’ current news consumption habits. In particular, phase 1.2 asked the following research questions.

- (RQ 1.2.1) Which device platforms do multilingual users prefer?
- (RQ 1.2.2) Which languages do multilingual users prefer per platform?

In order to answer these research questions, a user-centered study in the form of a semi-structured interview was conducted. The following sections first describe the experimental setup, followed by the study results.

3.2.2 Experiment Setup

The interview consisted of a total of 12 questions, which were grouped into three parts: “About you”, “About Your checking/browsing/reading multilingual news behavior” and “About Improvement for existing news access methods”. Six participants were interviewed in person for this study, and the average talking time was 40 minutes per person.

The workflow was strictly following the typical five interview stages, including “Ice breaker”, “Introduction”, “Key”, “Summary”, and “Wrap-up”.

Given that the focus of the study was on news checking, browsing, and reading behavior, participants were asked to indicate their different language writing and reading abilities. In particular, this self-rated proficiency ranged from 1 to 5, with 1 being the lowest proficiency, and 5 being the highest proficiency, as defined by the National Language Proficiency Definition13.

---

3.2.3 Study Results

3.2.3.1 Language abilities and input sources

Out of the six participants, four indicated knowledge of two languages, one could use three languages, and one could use four languages. In addition, we asked during the interview “How many language input sources are you using on your personal computer?”. Results showed that, for the users who had indicated knowledge of two languages, the same two languages had been installed as input sources. However, for participants with three languages or four languages, they indicated only having installed two or three language input sources on their devices respectively. Another interesting finding was that, no matter how many languages participants had proficiency in, only two languages were mentioned as regular news checking, browsing, and reading languages.

3.2.3.2 News Checking

With the growth of technology, there are more and more media types for checking news. Two main categories were defined for this study, namely “traditional media methods” and “new media methods”. Traditional media methods include TV, Radio, Newspaper, and Magazine. The new media methods are Search Engine, General News application or website, Professional News application or website, and Social Media. In order to obtain quantitative results for the analysis, we provided three questions for each news checking method, including “What’s your preference score for this method?”, “How often are you using it, Hourly, Daily, Weekly, or Monthly?”, and “Which language are you using?”.

Table 4 shows the interview results for each news checking media type. In the table, the preference score is the median score of all participants, the use frequency indicates the option(s) with majority votes, and the language-using result shows the option that obtained the majority votes.

As can be seen in Table 4, new media methods have a higher preference score than traditional media methods. All the participants responded they use search engines to follow hot news and to get in-depth news for gaining a better understanding. One participant said that she checks general news app or website when she remembers, and she uses email to subscribe to these websites. As for social networks, one participant mentioned that “you can see the ongoing updates from all the friends and channels, which is good for finding something different that they usually do not see elsewhere”. Additionally, many participants indicated that online social networks nowadays are becoming the most popular sites for news.
Table 4 Check News Methods

<table>
<thead>
<tr>
<th>News media type</th>
<th>News media name</th>
<th>Preference (1-dislike; 5-like a lot)</th>
<th>Use Frequent (H-Hourly, D-Daily, W-Weekly, M-Monthly)</th>
<th>Use which language? (L1-native language; L2-2nd language)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>TV</td>
<td>3.5</td>
<td>D, M</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Radio</td>
<td>3.5</td>
<td>D, M</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Newspaper</td>
<td>1</td>
<td>M</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Magazine</td>
<td>1</td>
<td>M</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Overall (Avg and majority vote)</td>
<td>2.25</td>
<td>M</td>
<td>L2</td>
</tr>
<tr>
<td>New</td>
<td>General search engine</td>
<td>3.5</td>
<td>D, W</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>General news app/website</td>
<td>4</td>
<td>D</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Professional news app</td>
<td>2.5</td>
<td>M</td>
<td>L2</td>
</tr>
<tr>
<td></td>
<td>Social network</td>
<td>4</td>
<td>D</td>
<td>Both</td>
</tr>
<tr>
<td></td>
<td>Overall (Avg and majority vote)</td>
<td>3.75</td>
<td>D</td>
<td>L2</td>
</tr>
</tbody>
</table>

Traditional media methods were found to have monthly usage frequency, whereas new media methods have a much higher frequency, which is daily. This matches the preference scores discussed in the previous paragraph, as participants prefer new media methods. For the traditional media methods, the use frequency is related to whether users have access to these methods. In particular, daily users of traditional media methods mentioned that it is mainly because they have the access to them, for example they have TV, whereas monthly users usually do not have access to them, for example they indicated not having a TV.

Most participants indicated high usage of English to check news, even though English is their second language. This is likely due to all the participants being international students staying in the U.S., and hence more English resources are available to them. However, an interesting observation was that there are many participants who use both languages for social media, because they care about both local news and the updates on social networks from their family and friends in their native country.

Some participants also mentioned how and why they are using these media methods. One participant said that she only reads a newspaper because the university provides a campus newspaper. Many of the participants used magazines for reading news when they are in
specific scenarios, such as waiting for a plane, on a plane, waiting in the waiting area of the shopping center, etc. And usually the magazines are provided in a public space.

### 3.2.3.3 News Browsing

Regarding news browsing behaviors, all participants responded that they choose news based on the title, keywords, and associated picture. Regarding how many results they want to see, participants mainly showed three types of patterns:

1. Based on number of news results: for example, one user mentioned that she usually browses 3-4 news articles, and only from the main page. Likewise, one participant said she only browses titles of 4-5 sources, and reads three articles.
2. Based on time: for example, one participant mentioned that she only reads news for about 10 minutes.
3. Depending on the media method: for the general news app, participants typically only read top news. However, for news feeds, participants would scan through an entire list. For social media, one participant mentioned he only scrolled down three times, whereas another participant said she would like to see all the latest updated news.

### 3.2.3.4 News Reading

Based on the news type, we divided the news into 11 categories, which includes Government, Society, Finance, Sport, Entertainment, Technology, Health, Religion, Education, Academic and Travel. Additionally, three main geo-locations are defined here as Local, National, and International.

To obtain the quantitative results for analysis, we provided one question for each news category, which is “Which language do you use for the different location types”. Table 5 shows the majority votes for language used for each news category.

As can be seen from Table 5, people tend to use both of their languages for local and national news. One reason is that some local news usually will not be published in other languages, and thus they use the language of the corresponding area to read those news. Another reason is that there might be a lack of news available in native language, due to government censoring, and therefore they need to read news in another language.
Table 5  News Categories and The Read Behavior

<table>
<thead>
<tr>
<th>News category</th>
<th>Use which language?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local</td>
</tr>
<tr>
<td>Government/Politics</td>
<td>Both</td>
</tr>
<tr>
<td>Society</td>
<td>Both</td>
</tr>
<tr>
<td>Finance/Business</td>
<td>Both</td>
</tr>
<tr>
<td>Sport</td>
<td>L2</td>
</tr>
<tr>
<td>Entertainment</td>
<td>Both</td>
</tr>
<tr>
<td>Technology</td>
<td>L2, Both</td>
</tr>
<tr>
<td>Health</td>
<td>Both</td>
</tr>
<tr>
<td>Religion/Spirituality</td>
<td>L2</td>
</tr>
<tr>
<td>Education/Collage</td>
<td>L2, Both</td>
</tr>
<tr>
<td>Academic</td>
<td>L2</td>
</tr>
<tr>
<td>Travel</td>
<td>L2, Both</td>
</tr>
</tbody>
</table>

For international news, participants tended to read in the native language or both languages. This is mostly due to international news being relatively major news, and thus they want to use another language to supplement what they read in their native language, e.g. “using English to read news from China”.

For most news categories, both languages were chosen. For example, in the health category, one participant mentioned she uses English to read news about “more up-to-date information like nutrition, calories”, but that she uses Chinese to read something about “how to balance the body based on ancient Chinese medicine theory”. However, for sport, technology and academic topics, most people choose L2, as the most complete and resourceful technology and academic news information were considered to be in English.

3.2.4 Summary and Current Impediments

The above interviews provide an initial analysis regarding current multilingual users’ news consumption behaviors and user preferences. Overall, the results have indicated that participants frequently read news in several languages, and that they see many advantages in doing so.

In general, participants thought that reading news in multiple languages can give them more complete information. For example, news websites were found to be typically biased, especially in the case of politics, and different languages were found to provide a more complete overview. Likewise, for some news stories, participants found that they are covered with more detail and/or depth in particular languages, whereas they are not covered or only briefly covered in other languages. It therefore seems that multilinguals like to check news
content from different languages to overcome information gaps, which may have been caused by simple information insufficiency in one language, or even by bias (as mentioned by several participants).

However, existing news access platforms have some limitations which currently prevent polyglots from easily finding and reading news content in different languages. First, while there is an abundance of news articles and news sources, users do not want to spend too much time on news selection and evaluation (i.e. they prefer to have an aggregate view that allows quick comparisons). Secondly, people often tend to do an initial search in their native language due to stronger language skills. However, in order to consume news in their second language, multilinguals need to manually conduct a separate search, which requires extra effort and may potentially be hampered by language skills. Likewise, when participants are reading news in a secondary language, they often encounter comprehension issues.

To improve the user experience of multilingual users, based on participants’ suggestions, an anticipated news platform therefore appears to be one that aggregates news content from multiple language news sources or channels (in order to allow comparisons), as well as one that shows top news for all languages when user searches for news in one language (i.e. not requiring an extra keyword input for each secondary language).

The results from this initial small-scale interview study have provided motivation for a larger-scale study (conducted in phase 2 – next chapter), which aims to gain more substantial evidence to better understand polyglots’ news consumption preferences, news querying and news results selection behaviours.
Chapter 4

Phase 2: An Investigation of Consumption, Querying, and Search Result Selection Behaviors

4.1 Introduction

As outlined in the introduction section, the objective of this thesis is to drive the development of novel multilingual news systems that better support the discovery and consumption of news in multiple languages, and that provide personalized support to a user’s individual abilities, preferences, and context. To develop such systems, it is imperative to first investigate current multilingual news access behaviors, as well as to perform an analysis of the different factors that may affect a user’s multilingual news consumption behavior. To this end, the research presented in this thesis chapter consists of two large-scale study phases performed using crowdsourcing platforms, which are aimed at investigating current general news consumption behaviors, as well as eliciting specific multilingual user behaviors in terms of news search querying and result selection. Specifically, phase 2 is aimed to build on the findings from phase 1.1. and 1.2. in the last chapter and extend the results through large-scale studies. The research questions associated with phases 2.1 and 2.2. are as follows:

Phase 2.1 consists of a large-scale survey investigating current general news consumption behaviors. There are four main research questions driving this study: (RQ2.1.1) Which platforms do multilingual users typically choose to search/browse news? (RQ2.1.2) Which topics do multilingual users typically choose to search/browse news? (RQ2.1.3) Which languages do multilingual users prefer for different news platforms and different news topics? (RQ1.4) How do multilingual users perceive the multilingual support provided by current systems?

Phase 2.2 consists of a series of large-scale task-based user studies that elicit multilingual news querying and result selection behaviors. There are three main research questions driving this study: (RQ2.2.1) Which of their language(s) do multilingual users generally use (and why) for querying as well as choosing news content? (RQ2.2.2) What is the effect of language proficiency on querying and/or result selection behaviors? (RQ2.2.3) What are the effects of news topic on querying and/or result selection behaviors?
4.2 Methodology

In order to address the above research questions, we conducted a series of studies on the Figure Eight\textsuperscript{14} crowdsourcing platform for both phases 2.1 and 2.2. This platform was chosen due to its high number of users and increasingly wide reach, providing access to participants across different geographical regions. For phase 2.1, we focused on the following languages: English, Chinese (Simplified), Chinese (Traditional), French, German, and Spanish. The focus on these languages was based on their significant online use (58\% of all Internet users\textsuperscript{15} are native speakers in at least one of these languages), and it ensured the collection of sufficient responses for each language, while at the same time representing a reasonable geographical spread. For phase 2.2, we focused on a subset of these languages, namely English, Chinese (Simplified), Chinese (Traditional), and Spanish (representing 52.8\% of all Internet users).

4.2.1 Phase 2.1: Multilingual News Consumption Behavior Survey

To study current multilingual news consumption behaviors, we conducted a survey on FigureEight that consisted of 12 questions (11 survey questions, 1 question for quality-control purposes), divided into 2 parts, namely 1) demographics and language proficiency, and 2) news consumption behavior. The specific questions are presented in Table 6.

Participants in this survey were required to have language proficiency in at least 2 of the 6 investigated languages. To get a well-distributed sample, 6 separate survey tasks were set up on FigureEight, one for each of the 6 different languages. Each survey collected responses from 200 participants, totaling 1200 participants for this phase overall.

4.2.2 Phase 2.2: Task-Based User Studies

The second phase investigated more specifically the language choices users make when issuing news search queries and selecting news search results. To this end, four separate task-based user studies were designed to investigate query formulation (Task-based Study 1), query selection (Task-based Study 2), and search result choices (Task-based Study 3, Task-based Study 4).

\textsuperscript{14}http://www.figure-eight.com (originally known as CrowdFlower)
\textsuperscript{15}http://www.internetworldstats.com/stats7.htm
<table>
<thead>
<tr>
<th>Categories</th>
<th>Question Id</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic and Language Proficiency</td>
<td>1</td>
<td>What is your age range?</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>What is your gender?</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>What is your country of origin?</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>What is your current country of residence?</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>How many of the languages mentioned in the instructions are you proficient in?</td>
</tr>
<tr>
<td>News Consumption Behavior</td>
<td>6</td>
<td>What platform(s) do you use for consuming news?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: TV and Radio; Paper-based newspapers and magazines; Search engine (e.g. Google, Bing, Baidu); Individual news apps/websites (e.g. CNN, Reuters); Aggregated news apps/websites (e.g. Apple News, JinriToutiao, Feedly, Flipboard); Social Media apps/websites (e.g. Facebook, Twitter, WeChat, Weibo).</td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>“Selected option in question 6” – How often do you consume news?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: Hourly; Daily; Weekly; Monthly.</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>“Selected option in question 6” – Use language type?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: Only first language; Only second language; Another language (Third or fourth language); More than one language.</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>(Quality check question) Please select the option that says, “I am multilingual”.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: I am not multilingual; I speak multiple languages; I am multilingual; I like speaking multiple languages.</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>For digital news, which device do you use?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: Desktop/laptop computer; Tablet; Mobile.</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Which news platform?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: Search Engine; General News Platform/APP; Professional News Platform/APP; and Social Media.</td>
</tr>
<tr>
<td></td>
<td>9.1</td>
<td>“Selected option in question 9” – Which Language you are using?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: Only first language; Only second language; Another language (Third or fourth language); More than one language.</td>
</tr>
<tr>
<td></td>
<td>9.2</td>
<td>“Selected option in question 9” – Which Language you use more/the most?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Options: First language; Second language; Another language (Third or fourth language); Equal.</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Have you experienced any issues when consuming news in different languages?</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Do you feel there is a lack in news availability in your first language?</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Do you have any other suggestion on how to improve the consumption of news in multiple languages?</td>
</tr>
</tbody>
</table>
The overall procedure for each of these studies was as follows: Participants were first asked to answer demographic questions and choose which 2 out of the 4 languages they are proficient in (English, Spanish, Chinese Simplified, Chinese Traditional – note however, that participants could not choose the combination of Chinese Simplified-Chinese Traditional). Participants were also asked to indicate their self-rated proficiency score (on a scale of 1-5). A set of 24 search topic descriptions was then presented sequentially to the participant. The specific task to be completed using these topic descriptions was slightly different for each study, as follows:

**Task-based Study 1:** This study focused on user query construction behaviors. Specifically, a participant was presented with a news topic description in each of their 2 languages (with the order of the topic description languages being counterbalanced), and then prompted to simply type in a query that they would enter into a news search engine to search for news on this topic (see Figure 12).

**Task-based Study 2:** This study concentrated on user query selection behaviors. A participant was again presented with topic descriptions in their 2 indicated languages. However, this study offered a dropdown list of possible queries, and participants were asked to choose one preferred query from this list (instead of providing a free text input field to type a search query as in task-based study 1). The drop-down list contained 2 queries (one query per user language, with language orders again being counterbalanced). Once a query was selected, a participant was prompted to indicate the reasons as to why he/she had picked this query (using a set of checkbox options and a free text input field), as shown in Figure 13.

**Task-based Study 3:** This study focused on user result list preferences. Similar to the above studies, participants were presented with a news topic description in each of their languages. In addition, the participants were presented with a pair of search result lists, with each list containing 10 results in a single language (i.e. one list containing results in language L1 and the other containing results in language L2, with L1 and L2 being the participant’s primary and secondary languages of proficiency). List ordering was counterbalanced to avoid any position bias. The user was then asked to indicate their preferred list. After the selection, a list of checkboxes appeared, prompting the user to indicate their reasons for selection (see Figure 14).

**Task-based Study 4:** This study aimed to gather search result list preferences similarly to study 3. However, the major difference in this study was that one of the two lists presented to
the participant contained results in the user’s primary language (L1) as the main body, and results in the user’s second language at the top of the list (called Mix-L1&L2 from here on). The other list, as well as the topic description and a predefined query were all in L1. Once a result list selection was made, the participant was presented with a list of checkboxes, and asked to specify why the participant had made the selection (see Figure 15).

Figure 12  Task-based Study 1 – writing a query in response to a topic description
Figure 13  Task-based Study 2 – choosing a query from a predefined list containing one query in L1 and one query in L2
Figure 14  Task-based Study 3 – choosing between an L1 list and an L2 list
Figure 15  Task-based Study 4 – choosing between an L1 list and a Mix-L1&L2 list
4.2.3 Participant Recruitment, Quality Control, and Payment
As previously mentioned, all participants were recruited through the FigureEight crowdsourcing platform. To minimize low quality entries, we took three quality control steps for all studies. The first step was to recruit only Tier 3 participants (the highest level on FigureEight). Secondly, the question choice options were shown randomly, and participants who always selected the same option number (e.g. always the first option) were removed. The last step consisted of a security check question that was added in the middle of the task questions. As previously mentioned, the question “Please select the option that says, ‘I am multilingual.’” was asked in the Phase 2.1 consumption survey. In phase 2.2, some of the tasks contained a check box stating “PLEASE CHECK THIS BOX (this confirms that you are not a robot or spammer)”. We manually filtered out the participants who selected the wrong options in the security check questions or showed obvious bias selection patterns. For the survey in phase 2.1, each participant was paid 15 cents per survey, with an average completion time of 4 minutes. In phase 2.2, each task-based study paid 25 cents per participant, and the average finishing time was 9 minutes.

4.2.4 News Topics, Translations, & Analysis
For Phase 2.1, the survey questions and choice options were generated based on a manual summarization of mainstream news platforms and news categories. This led to the following 6 categories of focus: ‘Government, Politics, and Events’, ‘Business and Finance’, ‘Health’, ‘Science and Technology’, ‘Sports’, and ‘Entertainment, Art, and Culture’. All questions and choice options were manually translated to support a total of six languages (English, Chinese-Simplified, Chinese-Traditional, French, German, and Spanish).

For Phase 2.2, to ensure unbiased/diverse sets of news topics, query choices, and result lists, a total of 48 topics were gathered from Google Trends16 and Baidu Index17 (12 topics per language * 4 languages), based on the largest population geolocation per language (i.e. 12 topics were gathered in English from Google Trends United States, 12 topics were gathered in Simplified Chinese from China Mainland, 12 topics were gathered in Traditional Chinese from Google Trends Taiwan, 12 topics were gathered in Spanish from Google Trends Mexico). To ensure quality translations, each news topic title (used as search query in studies 2, 3, and 4) and description was manually translated into the other 3 languages. Result lists

17https://index.baidu.com
were crawled using the Microsoft Bing Web Search API\(^\text{18}\) (10 per news topic). In total, there were 1600 participants (4 studies * 4 languages * 100 participants) for this phase.

All results were analyzed using Chi-squared tests, and each test used an alpha level of 0.05.

### 4.3 Results

#### 4.3.1 Phase 2.1: Consumption Survey

**4.3.1.1 Participant Demographics**

A total of 1200 participants took part in the multilingual news consumption survey, of which 1041 valid participants were retained after quality filtering as described in the previous section. Based on the requirements from this study, all participants had some proficiency (i.e. some reading/writing ability) in at least two of the 6 languages targeted in the study. Moreover, 215 participants were proficient in 3 languages, and 29 participants in 4 languages. There were 61.77% male participants, 35.35% female and 2.88% other (Figure 16). The vast majority participant age group (Figure 17) were ‘25-34’ (40.34%), followed by ‘35-44’ (23.44%), ‘18-24’ (20.75%), ‘45-54’ (9.7%), ‘55-64’ (4.71%), and ‘65+’ (1.06%).

![Participant gender percentage per language](http://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/)

**Figure 16** Participant gender percentage per language

In terms of participants’ L1 (i.e. their first language), ‘Spanish’ had the highest participant population with 33.62%, followed by ‘English’ 24.02%, ‘Chinese-simplified’ 15.47%, ‘German’ 11.53%, ‘Chinese-Traditional’ 8.93%, and ‘French’ 6.43%. In terms of L2 (i.e. their second language), ‘English’ had the largest participant group (55.24%), followed by ‘Spanish’ (12.97%), ‘French’ (10.95%), ‘Chinese-Traditional’ (10.66%), ‘Chinese-Simplified’ (6.05%), and ‘German’ (4.13%).

‘Desktop/laptop computer’ (57.64%) was stated as the most popular modality selection with statistical significance (chisq p-value<0.001), followed by ‘Mobile’ (31.32%), and ‘Tablet’ (11.04%).

4.3.1.2 News Platform Preferences (RQ2.1.1 & RQ2.1.3)

When asked which platforms they generally use to check news, participants most frequently indicated ‘TV/Radio’ (25.99%), followed by ‘Search Engine’ (20.52%), ‘Social Media’ (17.42%), ‘Individual News App/Web’ (13.81%), ‘Newspaper’ (13.62%), and ‘Aggregated News App/Web’ (8.65%). As shown in Figure 18, the most common news consumption frequency was “Daily”, followed by “Hourly”, “Weekly”, and “Monthly” (chisq p-value<0.001 within each platform, chisq p-value<0.001 for all platforms and usage choices). The only exception was noticed for ‘Newspaper’, where “Weekly” usage was indicated more than “Hourly” usage.

In addition to general usage frequencies per platform, we also enquired about general and specific frequencies with respect to different languages (see Figure 19). In general, the most
popular language choices were “Only 1st language” and “More than one language” (chisq p-value<0.001 within each platform, chisq p-value<0.001 for all platforms and language frequencies), followed by “Only 2nd language” and “Another language (3rd and 4th language)”. Interestingly, traditional news platforms, i.e. ‘TV/Radio’ and ‘Newspaper’, were noted to be used in “Only 1st language” much more than “More than one language”. By contrast, all digital methods (i.e. ‘Search Engine’, ‘Social Media’, ‘Individual News App/Web’, and ‘Aggregated News App/Web’) are used in “More than one language” more often than “Only 1st language”.

![Figure 18 Consumption survey platform time percentage](image1)

![Figure 19 Consumption survey platform language percentage](image2)
4.3.1.3 News Topic Preferences (RQ2.1.2 & RQ2.1.3)

In terms of general news topic category consumption frequencies, the most popular categories were ‘Government’ (23.15%), followed by ‘Science’ (18.75%), ‘Sports’ (15.69%), ‘Entertainment’ (15.21%), ‘Business’ (14.46%), and ‘Health’ (12.75%) (chisq p-value<0.001).

The vast majority of participants indicated that they currently use “Only 1st language” or “More than one language” most often across all topic categories (chisq p-value<0.001 within each topic) (see Figure 20), followed by “Only 2nd language” and “Another language (3rd and 4th language)”. Most of the topic categories had a higher indicated frequency for “Only 1st language” compared to “More than one language”, including for ‘Government’, ‘Sports’, ‘Business’ and ‘Health’. However, ‘Science’ and ‘Entertainment’ was shown to have an opposite usage pattern, with higher frequencies for “More than one language” compared to “Only 1st language”.

For participants who indicated using “More than one language”, we also asked specifically which language they used most often (or if languages were used to equal extents). Participants generally preferred to use their “1st language” over their “2nd language” and “Another language (3rd and 4th language)”, however, “Equal” language usage accounted for a large proportion of participants (see Figure 21).

![Figure 20 Consumption survey news topics language percentage](image)
4.3.1.4 Current System Support (RQ2.1.4)

The final questions of the survey enquired about participants’ subjective perception of issues with current news systems. To this end, we asked the following open-ended questions: “Q10. Have you experienced any issues when consuming news in different languages? [open-ended text box]”, “Q11. Do you feel there is a lack in news availability in your first language suggestions? [open-ended text box]”, and “Q12. Do you have any other suggestion on how to improve the consumption of news in multiple languages? [open-ended text box]”

Figure 22 presents an overview of answers to questions Q10 and Q11. As shown in this figure, while the majority of participants overall did not perceive issues, there were significant differences between languages. In particular, Chinese-Simplified participants showed a relatively high percentage (49%/44.7%) of “Yes” for questions Q10/Q11. For Q11, a large portion of Spanish users (35.5%) also indicated that their first language lacks available resources. Also, while a large proportion of participants technically mentioned that they did not experience any issues with current systems, this may have stemmed from the fact that they have become used to the monolingual nature of news platforms, and they have adapted their behavior accordingly. In fact, this was frequently mentioned in their open-ended responses (discussed below), and, as shown in phase 2.2, the use of multiple languages may indeed be increased significantly if supported more directly.
Figure 22 Answers to Q10. “Have you experienced any issues when consuming news in different languages?” and Q11. “Do you feel there is a lack in news availability in your first language suggestions?”

In addition to the quantitative analysis of questions Q10 and Q11, we also summarized participants’ open text responses for Q10, Q11, and Q12. After manual coding and summarization, there were a total of 775 clear and valuable answers for the three open-ended questions (197 for Q10, 224 for Q11, and 351 for Q12). One of the most frequently mentioned sets of issues when consuming news in multiple languages (Q10) related to ‘understanding’ content. This included general issues with understanding content as a whole, individual words, detailed parts of the content, advanced sentence structure, grammar, orthography, deep meanings, as well as differences in cultural backgrounds and different language expressions. Many of the participants also indicated that they had issues with the ‘News content’ itself (irrespective of their own language skills), including fake and biased news, as well as censorship. Another important set of issues related to ‘News availability’, including a general lack of L2 news, a specific lack of L2 news for some topic categories, not knowing where to find L2 news, or low update frequencies for L2 news. For many of the above issues, participants mentioned the need to spend ‘extra effort’ when trying to consume content in multiple languages, including having to use translators, having to switch searching/reading systems/approaches, and lacking an aggregation of different languages. By contrast, many participants also explained the reasons for indicating that there were “no issues” regarding consuming different language news (i.e. answering ‘no’ to Q10). The top
reason was that they thought there was always enough English news content available (when English was their L1 or L2), and that they knew where to find different language news sources (albeit requiring extra effort). Also, some participants indicated having good L1 and L2 proficiency, while others mentioned that they could use translators when encountering problems. One participant in particular noted that even though his/her L2 news consumption was slow, he/she could “tolerate” it.

When asked specifically about L1 news availability (Q11), both English L1 and Non-English L1 complained that there were issues regarding sparsity (needing to “waste time”, needing to “search more sites”, “hard to find trustworthy sources” and “requiring [manual] aggregation”). English L1 participants also frequently mentioned that there was often more “depth” to the news in their L2. Besides general (lack of) availability, Non-English L1 participants indicated that there were some issues with the news content and platforms themselves. Specifically, non-English language news was found to be less international, and not covering all topics. Additionally, non-English language content was found to lack details, originality, quality, professionalism, transparency, credibility/trustworthiness, objectivity, and political independence.

Lastly, participants provided an array of interesting and useful suggestions on how to improve the consumption of news in multiple languages (Q12). Most of the ideas related to ‘additional news system features’, especially adding translation functions such as page translators, sentence/word translators, or automatic translation upon mouse hovering (“A browser tool for hovering over words that would immediately give the reader a definition or related background information would be helpful.”). Additional suggestions related to the accuracy of the translator systems themselves. A number of participants also mentioned the development of multilingual news systems, which should aggregate news from different languages for easier comparison. Likewise, it was mentioned systems should provide article translations alongside links to the original. Furthermore, participants suggested the automatic summarization of multiple viewpoints and languages/regions.

### 4.3.2 Phase 2.2: Task-based User Studies

In addition to the above general news consumption survey, which was self-reported and subjective in nature, phase 2.2 consisted of a series of task-based user studies that specifically elicited user language choices when issuing news search queries and selecting search results.
As described in methodology section, there were 4 task-based user studies, with 4 different language versions per study (English, Spanish, Chinese Simplified, and Chinese Traditional).

### 4.3.2.1 Participant Demographics

In total, 16 studies (4 studies * 4 languages) were conducted on the FigureEight crowdsourcing platform, with each study having 100 participants. Thus, a total of 1600 participants took part in the task-based user studies, of which responses were retained from 1163 participants (359 for Task-based Study 1, 274 for Task-based Study 2, 260 for Task-based Study 3, and 270 for Task-based Study 4), after filtering out invalid entries using the quality controls described above.

In accordance with the guidelines of the study, all of the participants had proficiency in at least 2 out of the 4 languages. Table 7 presents a distribution of participants’ first (L1) and second (L2) languages.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>L1 (first/primary language)</th>
<th>L2 (second language)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task-based study 1</td>
<td>Spanish (31.11%), Chinese_Simplified (29.17%), Chinese_Traditional (21.11%), English (18.61%).</td>
<td>English (81.39%), Spanish (15.56%), Chinese_Traditional (1.94%), Chinese_Simplified (1.11%).</td>
</tr>
<tr>
<td>Task-based study 2</td>
<td>Spanish (33.94%), Chinese_Simplified (25.55%), Chinese_Traditional (21.9%), English (18.61%).</td>
<td>English (81.39%), Spanish (16.79%), Chinese_Traditional (1.46%), Chinese_Simplified (0.36%).</td>
</tr>
<tr>
<td>Task-based study 3</td>
<td>Spanish (32.31%), Chinese_Simplified (25.38%), Chinese_Traditional (23.46%), English (18.85%).</td>
<td>English (81.15%), Spanish (16.92%), Chinese_Traditional (1.15%), Chinese_Simplified (0.77%).</td>
</tr>
<tr>
<td>Task-based study 4</td>
<td>Spanish (34.07%), Chinese_Traditional (23.7%), Chinese_Simplified (23.33%), English (18.89%).</td>
<td>English (81.11%), Spanish (17.41%), Chinese_Traditional (1.48%), Chinese_Simplified (0%).</td>
</tr>
</tbody>
</table>

### 4.3.2.2 General Language Use (RQ2.2.1)

The first research question addressed by the task-based user studies was concerned with multilingual users’ general use of their languages when querying and selecting search results. As described in methodology section, Task-based study 1 asked participants to write their own queries in response to a topic description, while Task-based study 2 provided users with pre-written queries to choose from. Task-based study 3 asked users to choose between two
monolingual search result lists for a given topic (i.e. users were asked to choose between an L1-only list and an L2-only list), while Task-based study 4 asked users to choose between one monolingual (specifically an L1-only list) and one multilingual result list (i.e. a list with a mix of L1 and L2 results).

As shown in Figure 23, L1 usage was generally higher than other languages (language use in each study had statistical significance, Chi-squared p-value<.001). However, L1 usage varied significantly across the different studies, with L1 usage being the highest in Task-based study 1, and lowest in task-based study 4. Conversely, the use of L2 or Mix-L1&L2 was highest for task-based study 4, and lowest for task-based study 1. When specifically examining Non-English L1 participants, it was found that L2/Mix-L1&L2 was significantly higher than for English-L1 users.

Figure 23 Language choices for Task-based studies 1-4

In studies 2, 3, and 4, we additionally asked users to indicate why they had made their particular language choice. Specifically, participants could choose from a number of predefined reasons, or indicate a different reason in an open text box.

In study 2, 62.17% of queries were selected in L1 (and conversely 37.83% in L2). For this study, the participants could choose from the predefined reasons stated in Table 8. Overall, ‘Comfortableness’ was the top reason for choosing a specific language, followed by ‘Local Relevance’, ‘Domain/Topic Language Preference’, ‘Content Availability’, ‘International’,
‘Trustworthiness’, and ‘Proficiency’ (see Figure 24). A difference could be observed between English L1 and Non-English L1 participants, as shown in Figure 25. Overall, the top three reasons for English L1 participants were ‘Comfortableness’, ‘Proficiency’, and ‘Content Availability’. For non-English L1 participants, the top reasons overall were ‘Comfortableness’, ‘Local Relevance’, and ‘Domain/Topic Language Preference’. More specifically, they chose L2 queries over L1 queries because of ‘International’ and ‘Trustworthiness’ reasons (with statistically significant difference - Chi-squared p-value<0.001), as well as ‘Content Availability’ (with marginal statistically significant difference - Chi-squared p-value<0.09). By contrast, they chose L1 due to ‘Proficiency’, ‘Comfortableness’, ‘Domain/Topic Language Preference’ and ‘Local Relevance’ reasons (Chi-squared p-value<0.001).

Figure 24  Task-based study 2 – most frequent language choice reasons overall
In study 3, L1 result lists were selected 62.79% of the time, whereas L2 lists were selected 37.21% of the time. For this study, participants could choose from the set of predefined reasons presented in Table 9 to indicate why they had chosen a result list in that particular language. ‘Proficiency’ was selected most often overall, followed by ‘Local Relevance’, ‘Domain/Topic Language Preference’, ‘General Relevance’, ‘International’ and ‘Trustworthiness’ (see Figure 26). Again, a difference could be observed between English L1

---

**Figure 25** Task-based study 2 – relative language selection reasons for English L1 vs. Non-English L1 participants

**Table 8** Study 2 Language Selection Reasons

<table>
<thead>
<tr>
<th>Option Categories</th>
<th>Option as displayed to participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proficiency</td>
<td>My proficiency (skill) in other languages is not good enough.</td>
</tr>
<tr>
<td>Content Availability</td>
<td>There is typically more content available in this language.</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>Content in this language is typically more reliable (trustworthy).</td>
</tr>
<tr>
<td>Comfortableness</td>
<td>I am more comfortable searching and browsing in this language.</td>
</tr>
<tr>
<td>Domain/Topic Language Preference</td>
<td>I personally prefer this language for this type of topic (subject).</td>
</tr>
<tr>
<td>Local Relevance</td>
<td>Content in this language is typically more relevant to the area I live in.</td>
</tr>
<tr>
<td>International</td>
<td>Content in this language typically provides an international perspective.</td>
</tr>
</tbody>
</table>
and Non-English L1 participants (see Figure 27). Specifically, English L1 participants selected reasons relatively evenly (except ‘Proficiency’, where a big difference between L1 and L2 reasons was observed), with L1 always dominating L2 (all Chi-squared p-value<0.05). After having selected an L1 list, they frequently selected ‘Proficiency’, ‘Local Relevance’, ‘General Relevance’, and ‘Domain/Topic Relevance’. By contrast, after having selected an L2 list, they frequently selected ‘International’ and ‘Trustworthiness’. For Non-English L1 participants, the most important reasons overall were ‘Proficiency’, ‘Local Relevance’ and ‘Domain/Topic Relevance’. Their L1 selection showed statistically significantly higher values for ‘Proficiency’, ‘Local Relevance’, ‘General Relevance’ (Chi-squared p-value<.001), as well as ‘Domain/Topic Relevance’ (Chi-squared p-value<.01). By contrast, L2 selections showed statistically significantly higher values for ‘International’ (Chi-squared p-value<.001) and ‘Trustworthiness’ (Chi-squared p-value<.001).

![Figure 26 - Task-based study 3 – most frequent language choice reasons overall](image)

Figure 26  Task-based study 3 – most frequent language choice reasons overall
Figure 27  Task-based study 3 – relative language selection reasons for English L1 vs. Non-English L1 participants

Table 9  Study 3 Language Selection Reasons

<table>
<thead>
<tr>
<th>Option shorthand notation</th>
<th>Option as displayed to participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proficiency</td>
<td>My proficiency (skill) in other languages is not good enough. My proficiency (skill) in this language is better.</td>
</tr>
<tr>
<td>Content Availability</td>
<td>There is typically more content available in this language.</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>The results in this list seemed more reliable (trustworthy).</td>
</tr>
<tr>
<td>International</td>
<td>The results in this list contained an international perspective.</td>
</tr>
<tr>
<td>Domain/Topic Relevance</td>
<td>The results in this list were more relevant for this type of topic (subject).</td>
</tr>
<tr>
<td>Local Relevance</td>
<td>Content in this language is typically more relevant to the area I live in.</td>
</tr>
<tr>
<td>General Relevance</td>
<td>The results in this list were generally more relevant.</td>
</tr>
</tbody>
</table>

In task-based study 4, L1 result lists were selected 55.84% of the time, and Mix-L1&L2 result lists were selected 44.16% of the time. For this study, participants could choose from the predefined reasons presented in P-value<0.0 to indicate why they had chosen a result list in either L1 or a Mix-L1&L2 list. Overall, ‘Domain/Topic Enough Information-L1’, ‘More Complete Overview-Mix-L1&L2’ and ‘Domain/Topic Language Preference-L1’ were the top three selected reasons (see Figure 28). For participants who had selected an L1 result list, the top three reasons were ‘Domain/Topic Enough Information-L1’, ‘Domain/Topic Language Preference-L1’ and ‘General Language Preference-L1’ (Chi-squared p-value<0.001). By
contrast, for participants who had selected Mix-L1&L2, the top three reasons were ‘More Complete Overview-Mix-L1&L2’, ‘Domain/Topic Relevance-L2’ and ‘International-L2’ (Chi-squared p-value<0.001). As with the previous studies, a difference could again be observed between English L1 and Non-English L1 participants (see Figure 29). Firstly, Non-English L1 participants selected Mix-L1&L2 lists much more frequently overall. Secondly, while both English L1 and Non-English L1 participants frequently indicated ‘More Complete Overview-Mix-L1&L2’ and ‘Domain/Topic Relevance-L2’, Non-English L1 participants also frequently mentioned ‘International-L2’ as a reason.

Figure 28  Task-based study 4 – most frequent language choice reasons overall, when choosing either an L1-only result list or a Mix-L1&L2 result list

Figure 29  Task-based study 4 – most frequent language choice reasons for English L1 and Non-English L1 participants
<table>
<thead>
<tr>
<th>Selected Language</th>
<th>Option shorthand notation</th>
<th>Option as displayed to participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Domain/Topic Enough Information-L1</td>
<td>The L1 results contained enough relevant information for this topic.</td>
</tr>
<tr>
<td></td>
<td>Domain/Topic Preference-L1</td>
<td>I prefer using only L1 for this type of topic.</td>
</tr>
<tr>
<td></td>
<td>General Preference-L1</td>
<td>I always prefer using only L1.</td>
</tr>
<tr>
<td></td>
<td>Not Enough Proficiency-L2</td>
<td>My proficiency/skill in L2 was not good enough to read the L2 results.</td>
</tr>
<tr>
<td></td>
<td>Same Information-L1&amp;L2</td>
<td>The results in L2 contained the same information as the results in L1. No new information was provided.</td>
</tr>
<tr>
<td></td>
<td>Domain/Topic Irrelevance-L2</td>
<td>The results in L2 were completely irrelevant / unrelated to the topic.</td>
</tr>
<tr>
<td></td>
<td>More Complete Overview-Mix-L1&amp;L2</td>
<td>Seeing results in both L1 and L2 provided a more complete overview.</td>
</tr>
<tr>
<td></td>
<td>Domain/Topic Relevance-L2</td>
<td>The results in L2 were relevant for this topic.</td>
</tr>
<tr>
<td></td>
<td>International-L2</td>
<td>The results in L2 provided an international perspective.</td>
</tr>
<tr>
<td></td>
<td>Different Perspective/View-L2</td>
<td>The results in L2 provided a different perspective/view.</td>
</tr>
<tr>
<td></td>
<td>Domain/Topic Preference-L2</td>
<td>I always use L2 for this type of topic.</td>
</tr>
<tr>
<td></td>
<td>Local Relevance-L2</td>
<td>The results in L2 were relevant to the area I live in.</td>
</tr>
<tr>
<td></td>
<td>Domain/Topic Irrelevance-L1</td>
<td>The results in L1 were completely irrelevant / unrelated to the topic.</td>
</tr>
</tbody>
</table>

4.3.2.3 Effects of Language Proficiency (RQ2.2)

In order to further research the effect of language proficiency on language choice/selection, we investigated the impact of participant L2 language proficiency on query/result list language selections. Overall, as shown in Figure 30, participants with higher L2 proficiency generally showed increased L2 or Mix-L1&L2 choices (statistically significant for all studies - chi-squared test p-value<0.001). However, this pattern was significantly more pronounced for studies 1 and 2 compared to studies 3 and 4, i.e. participants with lower proficiency values still chose L2 or Mix-L1&L2 lists to a large extent in studies 3 and 4. Likewise, the pattern was more pronounced for English L1 participants compared to Non-English L1 participants, i.e. Non-English participants frequently chose L2 or Mix-L1&L2 queries/result lists regardless of proficiency.
Lastly, in addition to the effect of language proficiency, we analyzed the effect of news domain/topic on language choices. In particular, the news tasks shown to participants belonged to one of the following six categories: ‘Health’, ‘Science and Technology’, ‘Government, Politics, and Events’, ‘Business and Finance’, ‘Entertainment, Art, and Culture’, and ‘Sports’. As shown in Figure 31, different effects could be observed for English L1 participants versus Non-English L1 participants.

Figure 30  Query/Result language choices depending on user L2 proficiency

4.3.2.4 Effects of News Domain/Topic (RQ2.3)

Lastly, in addition to the effect of language proficiency, we analyzed the effect of news domain/topic on language choices. In particular, the news tasks shown to participants belonged to one of the following six categories: ‘Health’, ‘Science and Technology’, ‘Government, Politics, and Events’, ‘Business and Finance’, ‘Entertainment, Art, and Culture’, and ‘Sports’. As shown in Figure 31, different effects could be observed for English L1 participants versus Non-English L1 participants.
Specifically, for English L1 participants, query/result language choices were similarly distributed across different topics (statistical significance was only found for task-based study 1 (Chi-squared test p-value<.01), and task-based study 2 (p-value<0.001)). For example, in task-based study 2, it could be observed that for ‘Government, Politics, and Events’, participants selected L2 more often than for other topics/domains.

By contrast, news domain/topic had a statistically significant effect on query/result language situation in all studies for Non-English L1 participants (task-based study 1: Chi-squared test p-value<.01, task-based study 2: p-value<.01, task-based study 3: p-value<.01, task-based study 4: p-value<.001). In particular, ‘Science and Technology’ L2 query writing or selection showed nearly equal frequencies as L1 in Task-based study 1 and task-based study 2. Likewise, in task-based study 4, Mix-L1&L2 result lists were chosen almost to the same extent as L1 result lists selection for ‘Health’, ‘Science and Technology’ and ‘Sports’ topics.

4.4 Summary and Discussion

Through the above two phases (phases 2.1 and 2.2) of large-scale user studies, we found a number of interesting usage and behavior patterns regarding multilingual news access. In the first phase, a multilingual news consumption survey uncovered multilingual users’ subjective usage patterns regarding current searching/browsing/reading of news, while the second phase focused on eliciting query and result selection behaviors through task-based user studies.

As outlined in section 3 research questions, four specific research questions were addressed for phase 1, namely (RQ2.1.1) Which platforms do multilingual users typically choose to search/browse news? (RQ2.1.2) Which topics do multilingual users typically choose to search/browse news? (RQ2.1.3) Which languages do multilingual users prefer with different news platforms and different news topics? (RQ2.1.4) How easy is it currently to receive news from a platform in different languages and what are current barriers (and suggestions for improvement) to multilingual information access?

For (RQ2.1.1), ‘Search Engine’ and ‘Social Media’ were found to be the most popular current news media platforms, which reinforces the importance of focusing multilingual news research specifically on digital platforms. For (RQ2.1.3), the results on which specific language is used for which platform were found to be statistically significant, with modern digital platforms (‘Search Engine’, ‘Social Media’, ‘Individual News App/Web’, and ‘Aggregated News App/Web’) receiving a much higher percentage of “More than one
language” compared to traditional news outlets (e.g. newspapers). This result confirms again that multilingual users prefer to use more than one language while using these platforms, even though such platforms currently do not (directly) support language switching/aggregation.

In terms of general news topic/domain preferences (RQ2.1.2), the most popular news topics are ‘Government’, followed by ‘Science’, ‘Sports’, ‘Entertainment’, ‘Business’ and ‘Health’. The majority of participants were shown to choose either “Only 1st language” or “More than one language” while they are searching for information on these topics (RQ2.1.3) compared to “Only 2nd language” or “Another language (3rd and 4th language)”. Most of the topics (e.g. ‘Government’, ‘Sports’, ‘Business’ and ‘Health’) have similar percentages for “Only 1st language” and “More than one language”, with the former generally being slightly higher. Most strikingly, for ‘Science’ and ‘Entertainment’, multilingual users tend to use “More than one language” significantly more often than “Only 1st language”. These results highlight the fact that for most news topics, multilingual news access support is important and needed, and especially for specific topics where polyglots tend to prefer the use of multiple languages. Moreover, multilingual users might potentially use their L2 for an even wider variety of topics if they had access to a system that directly supported multilingualism (as tested and proven in phase 2.2)

Problems that participants claim to encounter while consuming news in multiple languages (RQ2.1.4) are often related to language understanding/proficiency reasons. Additionally, non-English L1 participants frequently mention that news content/availability is a problem in their L1, while English L1 participants do not experience this issue as often, again confirming prior work on language imbalance. This confirms again the content imbalance found by prior work and suggests that Non-English L1 participants might benefit most from multilingual news systems. Finally, suggestions on how to improve the consumption of news in multiple languages include the development of platforms/search engines that could retrieve multilingual news (based on the user’s language preferences), in order to compare and/or switch easily between languages, and the “depiction of multiple viewpoints”, as well as “comment sentiment summarization from different regions”.

Overall, phase 2.1 has thereby revealed that multilingual users frequently use multiple languages when looking for news information (particularly on digital platforms), and that individual users’ languages are used to different extents depending on the news topic domain.
In addition, multilingual users currently face a number of issues when trying to receive news in different languages, which often leads to a lack of complete and/or balanced news access, and these users would welcome new multilingual solutions that provide better support for multilingual news access.

While phase 2.1 had focused on subjective experiences with current news platforms, phase 2.2 focused on eliciting behaviors through task-based experiments (i.e. trying to elicit how users behave when given specific tasks). In particular, phase 2.2 investigated the following three specific research questions: (RQ2.2.1) Which of their language(s) do multilingual users generally use (and why) when querying and choosing search results? (RQ2.2.2) What is the effect of language proficiency on multilingual user query constructing/selection or result selection behavior? (RQ2.2.3) What are the effects of news domain/topic on multilingual user result selection behavior? To answer these questions, four separate series of task-based studies were set up, namely multilingual query writing (Task-based study 1), pre-written query selection (Task-based study 2), single language result list selection (Task-based study 3), and single language versus mixed language result list selection (Task-based study 4).

Overall, we found that both English L1 and Non-English L1 users tend to use their second language to a significant degree when writing and selecting news queries and result lists. This trend was increasingly the case with the later studies, i.e. users tended to use their L2 slightly less when they have to write their own queries, while they use their L2 significantly more when selecting from pre-written queries and in particular when choosing between different result lists (RQ2.2.1). Especially for Non-English L1 participants, L2 result lists (or Mixed-L1&L2 result lists) are chosen significantly more in comparison with English L1 participants. Most strikingly, for Task-based study 4, Non-English L1 participants chose L2 or Mix-L1&L2 result lists almost as often as L1 lists. These results provide strong evidence that polyglot users indeed wish to see news in more than just their first language, and that this desire should be addressed with novel multilingual news platforms that specifically support multiple language skills. In comparison with the results found in phase 2.1, the task-based user studies therefore provide strong evidence that multilingual users make much greater use of each of their languages when provided with a system that specifically supports multilingual searching (as opposed to the current monolingual paradigm).

In terms of reasons given for choosing specific languages, there are a number of differences between English L1 and Non-English L1 users. In particular, English L1 users
frequently mentioned that they used their L1 because of ‘Content Availability’, further stressing the content imbalance revealed in prior work and in phase 2.1. Likewise, Non-English L1 participants used their L2 (with L2 being English most of the time) for ‘Content Availability’. When given the option of accessing a multilingual result list (the Mix-L1&L2 result lists in task-based study 4), both English L1 and Non-English L1 participants frequently mentioned that they would choose the Mix-L1&L2 result list option because of ‘More Complete Overview’ and ‘Domain/Topic Relevance’. Additionally, Non-English L1 participants also chose Mix-L1&L2 lists for its additional ‘International’ value. It therefore appears that all multilingual users appreciate the additional value provided by a system that provides aggregations of news information from multiple languages, and that this alleviates some of the current system shortcomings mentioned in phase 2.1. Also, confirming some of the results of phase 2.1, the fact that Non-English L1 participants found L2 results to be more ‘International’ highlights the importance of English as an international language, especially for news (since most Non-English L1 participants’ L2 was English).

One important barrier to access that was found in phase 2.2 was L2 proficiency (similar to the results in phase 2.1), as it was frequently mentioned in both Task-based studies 2 and 3 as a reason for choosing L1 queries/result lists. This was also backed up by a specific analysis of the effect of proficiency (RQ2.2.2) on language choices, which found that participants with higher L2 proficiency tend to use or select L2 queries/lists much more frequently. In fact, the effect of language proficiency was most evident in study 1, where participants had to write their own queries. However, when being able to select from pre-written queries (study 2) or choosing between result lists (studies 3 and 4), this effect was not as pronounced. Especially task-based study 4 showed that, when given the option, multilingual users often choose news result lists that include L2 results, even if their proficiency is low. This lies in contrast with results from phase 2.1, where proficiency was mentioned as one of the most important barriers for using L2 at all. It therefore confirms that a system that directly supports multilingualism (either by providing automatic query translations and/or multilingual result lists) further encourages multilingual behaviors, and partially removes the hindrances of low language proficiency.

Lastly, in terms of the effect of news topic categories (RQ2.2.3), the results from phase 2.1 were largely confirmed, since L1/L2 usage in phase 2.2 also depended on the specific domain/topic. However, compared to phase 2.1, it was not only the ‘Science’ and ‘Entertainment’ news categories that led to increased L2 usage, but also a number of
additional categories (e.g. including ‘Government, Politics, and Events’, ‘Health’, and ‘Sports’). This confirms that multilingual users increase their use of multiple languages when provided with a system that directly supports multilingual news access (e.g. through providing mixed-language news results lists).

4.5 Conclusion
To study novel solutions for multilingual news access, this thesis chapter has presented the first large-scale investigation of multilingual news access behaviors. This investigation included two phases of user studies, which sought to survey current news consumption behaviors (phase 2.1), as well as news query and result selection behaviors (phase 2.2). The overall conclusions from phase 1 are that users currently already use multiple languages when looking for news online, and that they do so particularly when using digital platforms. It was also found that different languages are used for different purposes (in terms of news topic domain), and that Non-English L1 participants in particular make significantly more use of their L2 language (which is most often English). However, it was also found that this multilingual behavior is currently not well supported by news access systems, and that significant burden is placed on the user to seek balanced and diverse news information. Therefore, the results from phase 2.1 provide motivation for the development of novel multilingual news systems that better support users’ multiple language skills (as presented in phases 3 and 4). In addition, the high usage of mobile platforms has further highlighted the need to not only focus the development of such systems for the desktop, but also for mobile devices. The main conclusions from phase 2.2 are that multilingual users make significantly more use each of their different languages when a system provides multilingual support, such as when the system does not rely on users having to write queries in non-L1 languages (as the user’s proficiency may prevent them from doing so), and especially when a news search result page aggregates news from different languages. Therefore, phase 2.2 has provided further motivation for the development of multilingual systems that directly support the use of multiple languages, and has provided several design guidelines for doing so, in particular automatic query translation, as well as aggregate result presentation. However, further studies are necessary to explore different types of systems and interfaces in order to determine the best designs (as studied in phases 3 and 4).
Chapter 5

Phase 3: Mobile Pilot Study

5.1 Introduction

As outlined in chapter 2, the purpose of phase 3 is to guide initial designs of mobile multilingual news interfaces, as it was found that there was a distinct lack of prior work in this area. Specifically, the three research questions for phase 3 are:

- (RQ3.1) What type(s) of mobile multilingual news search interface designs are preferred by polyglot users?
- (RQ3.2) What are the interaction behaviors of polyglot users with mobile multilingual news search systems?
- (RQ3.3) Do different mobile multilingual search interfaces and/or news topics lead to different user behaviors?

In order to answer these research questions, as well as to find the most appropriate mobile interface designs to use in a larger study in phase 4, this chapter conducts a pilot comparative analysis of user behaviors and preferences with respect to 4 (2 pairs) different multilingual news mobile search interfaces, through a pilot lab study involving 10 participants.

The chapter is structured as follows. Section 5.2 first outlines the experimental setup for the study, followed by the data analysis results in section 5.3. Finally, section 5.4 summarizes the chapter with a discussion of the evaluation results.

5.2 Experimental Setup

This section first describes in detail the 4 different mobile interfaces that were designed and developed for this pilot study. This is followed by a description of the experimental platform, study tasks, user study, and analysis procedure.

5.2.1 Interface Used in the Study

The interfaces used in phase 3 were based on the techniques identified from prior work on desktop interfaces, and particularly the interactive multilingual interfaces presented in phase 1 (namely ‘Panels’, ‘Universal’, ‘Tabbed’, and ‘Interleaved’). However, since ‘Interleaved’ had been found to be strongly disliked overall in phase 1, this design was not included in
phase 3. Likewise, due to the limited screen size on mobile devices, the ‘Panels’ interface was also not studied in phase 3. Therefore, phase 3 only included variations of the ‘Universal’ and ‘Tabbed’ interface types. Specifically, two variations each were designed for ‘Universal’ and ‘Tabbed’.

‘Universal’ Interface variations: ‘CollapseAll’ and ‘CollapsePartial’. Both of these interfaces present groups of vertical blocks for different language results in a single list. For example, as shown in the ‘CollapseAll’ example in Figure 32, when a user selects two languages, a block of 6 results in the user’s first language is shown above a block of six results in the user’s second language. In addition, if a user clicks/taps the language bar with the “-” sign (collapse sign) on the right, the corresponding result section collapses. The “-” sign then turns to a “+” sign (spread sign), and once this is clicked/tapped again, the whole block will spread again. By contrast, the ‘CollapsePartial’ interface (Figure 33) initially only shows three results per language, thereby allowing users to see an equal amount of results per language above the fold. If users would like to see additional results for a particular language, they can click the “+” button to show a further 3 results (and later click the “-” sign to collapse down to three again).

Figure 32  CollapseAll Interface

Figure 33  CollapsePartial Interface
‘Tabbed’ Interface Variations: ‘Slide’ and ‘Tabbed’. This pair of mobile search interfaces are both inspired by the design of the desktop ‘Tabbed’ interface presented in phase 1. In particular, the mobile ‘Tabbed’ interface design (Figure 34) closely follows the desktop version, whereby users can change between lists of results in different languages through clicking tabs. While the ‘Slide’ interface (Figure 35) also has the same single language list per page design, the switch between languages is performed through a sliding function invoked by tapping the green arrows (as is commonly done in mobile interfaces).

5.2.2 Platform, Focus, and Participants

To test these 4 different mobile interfaces, we conducted a lab-based study using a purpose-built experimental platform (built using PHP and MySQL), running on a mobile device with a 5.5-inch screen.

The main focus of the study was to perform two comparative evaluations, namely a comparison between ‘CollapseAll’ and ‘CollapsePartial’ (the two variations of the ‘Universal’ desktop search design type), and a comparison between ‘Slide’ and ‘Tabbed’ (the two variations of the ‘Tabbed’ desktop search design type). This focus was to narrow down the
total number of interfaces to two (one per design type), in order to be able to conduct a larger study in phase 4 (which also included desktop versions).

Experiment participants were recruited through campus mailing lists and were required to have 'some proficiency’ in at least two out of a set of 4 languages (English, Spanish, Simplified Chinese, and Traditional Chinese). Each participant was given a gift voucher worth 5 USD, and the average time to complete the entire study was approximately 20 minutes.

5.2.3 Task Questions

Each user was presented with a total of 16 news topic tasks (P-value<0.1), which were to be completed using the 4 interfaces (1 practice task per interface, 3 real tasks per interface, counterbalanced across participants/interfaces/news categories). The news topics were generated based on a manual summarization of mainstream news platforms and news categories. This led to the following 6 categories of focus: ‘Government, Politics, and Events’, ‘Business and Finance’, ‘Health’, ‘Science and Technology’, ‘Sports’, and ‘Entertainment, Art, and Culture’. All questions and choice options were manually translated to support a total of four languages (English, Chinese-Simplified, Chinese-Traditional, and Spanish). The focus on these languages was based on their significant online use (52.8% of all Internet users are native speakers in at least one of these languages), and it ensured the collection of sufficient responses for each language.

<table>
<thead>
<tr>
<th>Task Type</th>
<th>News Categories</th>
<th>News Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Sports</td>
<td>World Cup</td>
</tr>
<tr>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>Storm Severe weather Tornado Damage</td>
</tr>
<tr>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>US North Korea Summit Nuclear</td>
</tr>
<tr>
<td>Real</td>
<td>Business and Finance</td>
<td>Melon Japan Fruit Yubari King</td>
</tr>
<tr>
<td>Real</td>
<td>Business and Finance</td>
<td>Starbucks Biased Diversity Training Close</td>
</tr>
<tr>
<td>Real</td>
<td>Health</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hot tub Swimming Pool</td>
</tr>
<tr>
<td>Real</td>
<td>Health</td>
<td>Hamburger Hot dog health</td>
</tr>
<tr>
<td>Real</td>
<td>Science and Technology</td>
<td>Lenovo Sony Xperia Z5 Smartphone</td>
</tr>
<tr>
<td>Real</td>
<td>Science and Technology</td>
<td>Microsoft GitHub acquire</td>
</tr>
<tr>
<td>Real</td>
<td>Sports</td>
<td>NCAA Basketball</td>
</tr>
<tr>
<td>Real</td>
<td>Sports</td>
<td>New York Yankees baseball</td>
</tr>
<tr>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Star Wars Episode IX</td>
</tr>
<tr>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Marriage Harry Potter Neville</td>
</tr>
</tbody>
</table>
The specific news topics were selected from the previously presented task-based user study in phase 2. In particular, to ensure unbiased/diverse sets of news topics, a total of 13 topics were gathered from Google Trends\textsuperscript{19} and Baidu Index\textsuperscript{20} (1 practice task, 12 real tasks). To ensure quality translations, each news topic and description was manually translated into the other 3 languages. Result lists were crawled using the Microsoft Bing Web Search API\textsuperscript{21}.

Before performing assigned tasks with each interface, participants first performed a practice task (the same practice task was used for each interface).

### 5.2.4 Procedure

Similar to the study presented in phase 1.1., there were four main steps each participant followed in the experiment. These five steps were i) completing an initial questionnaire, ii) performing tasks, iii) completing mid-study questionnaires, iv) completing mid-post-study questionnaires, and v) completing a post-study questionnaire.

The first step was a demographic questionnaire that included self-reported language proficiency in terms of reading, listening, and writing ability. Proficiency was measured on a scale from 1 to 4, with one corresponding to very limited proficiency, and 4 corresponding to native proficiency.

Next, users interacted with the previously described interfaces. More specifically, for each interface, users performed one practice task, followed by 3 tasks. For each task, participants were asked to issue their own queries to look for information on the given news topics, with the topic being displayed in all of the participant’s languages. Participants were also asked to mark any results they found relevant to the news topic content, using buttons placed next to each result. These markings were recorded by the experimental platform, along with task times, number of queries, as well as number of viewed documents (i.e. news results clicked/opened). Participants were free to move on to the next task if they felt that they had received sufficient relevant information on the task. Upon completion of all 3 tasks with one interface, a mid-study questionnaire was presented, which asked users to indicate their perceived level of search support provided by the interface.

\textsuperscript{19}http://trends.google.com/trends/trendingsearches/realtime?geo=US&category=all
\textsuperscript{20}https://index.baidu.com
\textsuperscript{21}http://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/
After having interacted with one pair of interfaces, a mid-post-study questionnaire asked participants to indicate their overall comparative impressions, including i) which interface was ‘easier to use’, ii) which interface they ‘preferred the most’, and iii) which interface they ‘disliked the most’. In addition, participants were asked to indicate their reasoning for particular preferences/dislikes.

Lastly, after having interacted with all interfaces, a post-study questionnaire asked participants to indicate their overall comparative impressions, including i) which interface was ‘easiest to use’, ii) which interface they ‘preferred the most’, and iii) which interface they ‘disliked the most’. In addition, participants were asked to indicate their reasoning for particular preferences/dislikes.

5.2.5 Data Analysis

All results were analyzed using Chi-squared tests, and each test used an alpha level of 0.05.

5.3 Results

5.3.1 Participant Demographics

A total of ten participants took part in the user study. Two countries of origin were reported by our participants, which are USA (80%) and China (20%). All participants were between 20 and 22. In terms of participants’ first language (i.e. the most proficient language-L1), the vast majority of participants (70%) indicated English as their first language, followed by Spanish (10%), Simplified Chinese (10%), and Traditional Chinese (10%) (Figure 36). For the secondary language (i.e. the second most proficient language-L2), the most common languages were Simplified Chinese (30%), English (30%), Spanish (20%) and Traditional Chinese (20%) (Figure 36).
Regarding language proficiency, since L1 was the native language, proficiency level for all participants was “5”. For L2 proficiency, there were great variations as shown in Figure 37. Most of the participants had level “4” (50%), followed by “2” (40%) and “3” (10%). All ten participants indicated that they use Web Search engines such as Google on a daily basis.

5.3.2 User Preferences

Since the purpose of phase 3 was to gauge initial user reactions regarding different design ideas, the main goal was not to perform a large-scale user study, but rather gain qualitative insights. It was therefore expected that a statistical analysis as in phase 1 may not result in any statistically significant results (due to sample size). Nonetheless, a statistical analysis was performed for completeness. As show in Figure 38, when doing pairwise comparisons between ‘CollapseAll’ and ‘CollapsePartial’, ‘CollapseAll’ seemed to be considered the ‘more preferred’ and ‘easier to use’ by most participants, while ‘CollapsePartial’ is the ‘more disliked’ (easiness p-value<0.21; preference or dislike p-value<0.53).
‘Slide and ‘Tabbed’, it was found that ‘Tabbed’ was generally considered the ‘more preferred’ and ‘easier to use’ by most participants, while ‘Slide’ was the ‘more disliked’ (easiness or preference or dislike p-value<0.53). When comparing all interfaces, ‘CollapseAll’ was generally considered the easiest interface, ‘Slide’ and ‘Tabbed’ are the most preferred interfaces (easiness p-value<0.76, preference p-value<0.95, dislike p-value<0.16), and, interestingly, ‘Slide’ was also the most disliked interface overall. No participants selected ‘Tabbed’ as their most disliked interface.

![Figure 38 Post-study questionnaire results](image)

In addition to these statistical analyses, a thorough investigation of participants’ qualitative answers was performed. Among participants’ comments about the easiness of ‘CollapseAll’ and ‘CollapsePartial’ interfaces, most of them indicated that ‘CollapseAll’ is easier to use because it provides an overview by letting user skim over the maximum amount of results at once, and it allows users to get a sense of information relevance without extra click actions, as well as it offers the collapse sections/buttons by allowing users to tap away or hide the specific language space that the users may not want to see. However, a small portion of participants found ‘CollapsePartial’ easier because they wanted to get an initial sense of different language results, before deciding if they had an interest to see more results from a single language instead of seeing too much information at the beginning. Regarding preferences, participants who preferred ‘CollapseAll’ mostly indicated that it is easier to see
all the information at once without clicking any additional buttons, hence it was considered faster/easier/simpler to use. Others motioned it has a more modern/cleaner/more beautiful design. However, some participants disliked ‘CollapseAll’, because they felt that too much information was presented at once, which did not allow them to compare the information provided from both languages (and closing an entire section in the ‘CollapseAll’ interface could be troublesome because there might have something the user still wanted to be able to see without having to click to reopen the section). On the other hand, participants who chose ‘CollapsePartial’ as the more preferred interface liked the ability to decide to see more particular language results or not, instead of the overwhelming information upfront. Though most of the participants disliked it due to the interface’s “see more results” button, for instance, they wanted to see more news articles at once without having to click to see more, or they did not like the position of the button.

When comparing the ‘Tabbed’ and ‘Slide’ interfaces, the majority of participants considered ‘Tabbed’ easier to use, since it was easier to see what languages were available, using the tab itself that showed which page/language the user is on (and they just needed to click on a new tab button to see another language). Some users, however, felt that the ‘Slide’ interface was easier to use because of the button for language switching access (i.e. ‘Slide’ buttons were “conveniently” located for users’ hand positions, whereas the tab buttons were harder to reach and felt uncomfortable for the user’s hands). Some participants also noted the “obviousness” of slide buttons, i.e. ‘Slide’ buttons are clear and visible when they are reading news, while the tabs can be easily missed/forgotten about. In addition to ‘easiness’, the ‘Tabbed’ interface was also preferred overall, because it was considered more intuitive, and it clearly indicated which language the user was reading/could select from. However, some reasons for disliking ‘Tabbed’ were the small size of tabs, hard to read tabs, users having to stretch their hands to reach the tab button, and, when users needed to switch languages, they needed to go back up to the top of the webpage. The ‘Slide’ was therefore preferred by some participants because they thought the language switching was more comfortable due to the more convenient slide buttons. However, ‘Slide’ was generally not liked overall because of several reasons: 1) large visual overlaps 2) dislike of the left/right swiping action 3) bad action mapping, as left-right action may also mean previous/next page.

When comparing all four interfaces, ‘CollapseAll’ was considered the easiest to use because users felt they could better see both languages at once, with clear language indication and without additional clicks. The ‘CollapsePartial’ interface was sometimes selected as most
preferred because it was found to have a clear layout for both languages, with the most recent and relevant news at the top without too much information overload. Also, the icons were found to be easier to notice. When ‘Tabbed’ was indicated as easiest to use, participants indicated their liking of having the maximum number of search results at once, and that they could immediately understand the method of toggling between languages without being visually distracted or confused. Reasons for liking ‘Slide’ included the display of all the relevant information for one language at a time, making it easier for some participants to understand. Also, it required the least amount of effort for switching languages since the user could simply tap a button that was considered easy to reach.

In terms of reason for selecting their most preferred interface, some participants thought that the ‘Tabbed’ interface has a clean and visually pleasing design, as well as easier to understand. It was also found to keep information more organized and let the participants see the information in the language they want to see without distraction. The ‘Slide’ interface was liked because the swap arrow design was considered pleasing and easy to use, letting participant switch between languages fast, as well as enabling the participant to see one language at a time. ‘CollapsePartial’ was considered clear and presenting the most relevant news for both languages well. Also, the “see more” icons were considered obvious to notice. Participants who preferred ‘CollapseAll’ complemented on its ease of use and aesthetics, and the fact that it allows them to see a lot of news at once while still indicating which language content they are looking at.

As previously mentioned, ‘Slide’ was the most disliked interface overall, followed by ‘CollapsePartial’ and ‘CollapseAll’, while no participants vote ‘Tabbed’ as most disliked. A key reason mentioned for ‘Slide’ was the design for swapping between languages, as it was considered confusing and inconvenience. Also, it does not allow participants to compare between two languages. One of the reasons mentioned by users who voted ‘CollapsePartial’ as most disliked is that ‘CollapsePartial’ displays insufficient amount of information in the first place and it requires extra efforts from users to click to see more results. On the contrary, participants did not like ‘CollapseAll’ because they think it presents too many results, which leads to information overload.

In addition to these overall comments, phase 3 also included a mid-survey questionnaire (after each individual interface had been interacted with), containing 5 questions (see figure 39). The ‘Tabbed’ interface was found to have the highest ratings for Q1 (enough information)
and Q4 (easily identify relevant information), while ‘CollapseAll’ had the highest ratings for Q3 (easily combine information) and Q5 (information overview). The ‘Slide’ interface had the highest average rating in Q2 (information variety). However, as with the above analyses for overall preferences, statistical analyses (using ANOVA) did not find a statistically significant effect on rating.

Figure 39 Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree).

5.3.3 General User Behavior

In terms of user behavior, it was found that there was a statistically significant difference in terms of task time (p-value<0.05). In particular, the ‘Slide’ and ‘Tabbed’ interfaces had the
fastest completion time, while ‘CollapseAll’ and ‘CollapsePartial’ were significantly slower. In addition, ‘CollapseAll’ was faster than ‘CollapsePartial’.

![Figure 40 Interface and Participant Per Interface Average Time](image)

### 5.4 Summary and Discussion

In this section, we summarize the key findings from our pilot study regarding mobile interfaces, specifically related to user preferences and interaction behaviors (RQ3.1 and RQ3.3), as well as general interaction behaviors found across all interfaces (RQ3.2). We also provide further insights and recommendations based on participants’ free text responses, and discuss implications for designing multilingual mobile search interfaces.

When comparing ‘CollapseAll’ and ‘CollapsePartial’, the ‘CollapseAll’ interface was significantly preferred and considered easier to use, and it was regarded as providing a better ‘overview of information’ to allow user to skim through the maximum amount of results at once without extra clicking. It therefore appears that the need to scroll to view the results for the second language was not an issue. In addition, when a user does not want to see the results for some languages, it allows the user to minimize (i.e., tap away and hide) the specific language space. However, several participants still reported some issues interacting with the ‘CollapseAll’ interface. Especially, some participants felt that it was overwhelming to have all the information presented at once, which made it harder to compare two languages’ search results. Some participants preferred the ‘CollapsePartial’ interface due to its option to control whether to display more particular language results after getting an initial sense of different languages’ search results at first. However, most of the participants disliked it...
because participants had to do additional clicks on the interface’s “see more results” button to see more results.

When comparing ‘Tabbed’ and ‘Slide’, the majority of participants considered the ‘Tabbed’ interface to be more comfortable to use and it was generally more preferred, because it was considered more intuitive and easier to view which language the participant was reading (or was available for selection). However, there were several other participants who disliked the ‘Tabbed’ interface, due to the tab button’s small size and position in the screen, for which participants have to stretch hands or fingers to reach or go back up to the top of the webpage (which can be easily missed or forgotten about). Some participants felt that the ‘Slide’ interface was therefore easier to use or more preferred because the buttons for language switching are clear and visible and close to the user’s finger position. However, ‘Slide’ was generally not preferred by participants because its button design has visual overlaps with the news content, and it might create confusing interpretation since left/right action may also mean previous/next page.

When comparing all four mobile interfaces, ‘CollapseAll’ interface was considered the easiest to use and the most preferred overall, because participants felt they could better read result lists of both languages at once, with its clear language indication and without additional clicks. The most controversial interface was ‘Slide’, as it was voted first as the ‘most disliked’ interface as well as the ‘most preferred’ interface. Its slide switch language button requires the least amount of effort from users for switching languages since the switch button is close to the user’s finger, but it still visually overlaps the news content or might confuse participants about its functionality.

In terms of general user interaction behaviors, one interesting finding across all mobile interfaces was that ‘Slide’ and ‘Tabbed’ interfaces had the shortest task completion time and they are both voted as the ‘most preferred’ interface. Also, ‘CollapseAll’ was more preferred and had shorter task completion time than ‘CollapsePartial’.
Chapter 6

Phase 4: Multilingual News Desktop and Mobile Comparison User Study

6.1 Introduction

Following the pilot study in the last chapter (phase 3 - focusing on mobile interface designs for multilingual news search), as well as the desktop study presented in chapter 3 (phase 1.1 – focusing on desktop interfaces for general multilingual Web search), the objective of phase 4 is to specifically evaluate different interactive multilingual news search interface designs, for both desktop and mobile. This is aimed to get a better understanding of how to best support polyglot Web users and their multiple language skills when searching for news in multiple languages, on both desktops and mobile devices. Specifically, we ask the following research questions:

• (RQ4.1) What type(s) of interactive multilingual news search interface designs are preferred by polyglot users? Specifically, what type(s) of interfaces are preferred on desktop and mobile devices?

• (RQ4.2) What are the interaction behaviors of polyglot users with different interactive multilingual news search interfaces, on the desktop and mobile devices?

• (RQ4.3) Do different interactive multilingual search interfaces and/or news topics lead to different user behaviors?

In order to answer these research questions, this thesis chapter provides the first in-depth comparative analyses of user behaviors and preferences with respect to five different multilingual search user interfaces across both desktop and mobile, through a lab-based user study involving 40 participants.

The remainder of this chapter is structured as follows. Section 6.2 first describes the experimental setup for the multilingual news search interfaces user preference and behavior study. Section 6.3 presents the data analysis results. Lastly, section 6.4 concludes the chapter with a discussion of the evaluation results.
6.2 Experimental Setup

As mentioned in the previous section, this phase focuses specifically on studying and comparing different types of interactive multilingual news search interfaces, on both desktop and mobile devices. This section first describes in detail the 5 different interfaces that were designed and developed for the user study. This is followed by a description of the experimental platform, study tasks, and user study and analysis procedure.

6.2.1 Interface Used in the Study

The interfaces presented in this section are based on the techniques identified from prior work, as well as the studies presented in phases 1.1. and 3. In particular, the desktop interfaces presented in this section are extending the interactive multilingual desktop interfaces presented in phase 1, while the mobile interfaces are a selection of the most preferred interfaces from phase 3. Specifically, the study includes desktop ‘Tabbed’ and mobile ‘Tabbed’ interfaces, as well as desktop ‘Universal’ and mobile ‘CollapseAll’ interfaces. In addition, the study includes the most preferred interface from phase 1, namely ‘Panels’ (which does not have an equivalent mobile version, due to screen space restrictions).

**Desktop Tabbed Interface.** As shown in Figure 41, the desktop ‘Tabbed’ interface is the same as in phase 1, which allows users to change between lists of results in different languages through tabs. The ‘Tabbed’ interface automatically translates (using the Microsoft Translator API\(^\text{22}\)) the user’s query to the language of the currently selected tab. Translations can also be edited. Each result page displays a single monolingual ranked list, which contains 12 results (retrieved using the Microsoft Bing Search API\(^\text{23}\)). At the bottom of the page, participants can click ‘previous’ and ‘next’ buttons to view more search results.

![Desktop Tabbed Interface](image)

**Figure 41 Desktop Tabbed Interface**

\(^{22}\)datamarket.azure.com/dataset/bing/microsofttranslator - While there are many different machine translation techniques and tools, this was not the focus of this study. The Microsoft Translator API supports a large number of language pairs, and it was found to work sufficiently well for the topics used in the study tasks.

\(^{23}\)datamarket.azure.com/dataset/bing/search
Desktop Panels Interface. The ‘Panels’ interface (Figure 42) displays one or more panels on a single result page, with each panel displaying results in a different language. The number of languages (and hence panels) per page is determined by the user through the use of checkboxes (up to 2 languages). A user may choose to view results in each of his/her languages, or only focus on one single language.

To make sure that the same number of results is displayed for each result page, and to ensure consistency across interfaces (e.g. ‘Panels’ having the same number of results per page as ‘Tabbed’), the number of results per language equals to 12 divided by the number of languages. For example, if a user has selected 2 languages, each individual panel will contain 6 results. Conversely, if only one language is chosen, the result list is equivalent to the ‘Tabbed’ interface, i.e. 12 results for the chosen language. Each panel is headed by the (editable) auto-translated query according to the respective language, and the results within each panel can be navigated with ‘previous’ and ‘next’ buttons.

![Desktop Panels Interface](image)

Figure 42 Desktop Panels Interface

Desktop Universal Interface. The ‘Universal’ search interface also presents a single result list similar to the ‘Panels’ interface. However, results from each language are grouped together into horizontal blocks, rather than vertical blocks. For example, as shown in Figure 43, when a user selects 2 languages, a block of 6 results in the user’s first language is shown above a block of 6 results in the user’s second language.
Mobile CollapseAll Interface. The ‘CollapseAll’ search interface is similar to the desktop ‘Universal’ interface, presenting groups of vertical blocks for different language results in a single list. For example, as shown in the ‘CollapseAll’ example in Figure 44, when a user selects two languages, a block of 6 results in the user’s first language is shown above a block of six results in the user’s second language. In addition, if a user clicks/taps the language bar with the “-” sign (collapse sign) on the right, the corresponding result section collapses. The “-” sign then turns to a “+” sign (spread sign), and once this is clicked/tapped again, the whole block will spread again.

Mobile Tabbed Interface. The ‘Tabbed’ search interface closely resembles the desktop ‘Tabbed’ interface, allowing users to change between lists of results in different languages through tabs. At the bottom of the page, participants can click ‘previous’ and ‘next’ buttons to view more search results.
Figure 44 Mobile CollapseAll Interface

Figure 45 Mobile Tabbed Interface
6.2.2 Platform and Participants

We conducted a lab-based study using a purpose-built experimental platform (built using PHP and MySQL), running on a desktop computer with a 24-inch monitor, as well as a mobile device with a 5.5-inch screen.

Experiment participants were recruited through campus mailing lists and were required to have ‘some proficiency’ in at least two out of a set of 4 languages (English, Spanish, Simplified Chinese, Traditional Chinese). Each participant was given a gift voucher worth 15 USD, and the average time to complete the entire study was approximately 45 minutes.

6.2.3 Task Questions

Each user was presented with a total of 17 news topics (P-value<0.2), which were to be completed using the 3 desktop interfaces and 2 mobile interfaces (1 practice task per interface, 3 real tasks per interface, counterbalanced across participants/interfaces/news categories). The news topic categories were generated based on a manual summarization of mainstream news platforms and news categories. This led to the following 6 categories of focus: ‘Government, Politics, and Events’, ‘Business and Finance’, ‘Health’, ‘Science and Technology’, ‘Sports’, and ‘Entertainment, Art, and Culture’. All questions and choice options were manually translated to support a total of four languages (English, Chinese-Simplified, Chinese-Traditional, and Spanish).

To ensure unbiased/diverse sets of news topics, a total of 26 topics were gathered from Google Trends\(^\text{24}\) and Baidu Index\(^\text{25}\) (2 practice tasks, 24 real tasks). To ensure quality translations, each news topic and description was manually translated into the other 3 languages. Result lists were crawled using the Microsoft Bing Web Search API\(^\text{26}\).

Before performing assigned tasks with each interface, participants first performed a practice task (the same practice task was used for each interface).

6.2.4 Procedure

The study was composed of two main stages: the first stage consisted of the “study stage”, which involved participants performing tasks with the above-mentioned desktop and mobile interfaces. Specifically, participants first interacted with all interfaces on one platform (either


\(^{25}\)https://index.baidu.com

\(^{26}\)http://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/
desktop or mobile), before moving on to using all interfaces in the other platform. To avoid any bias, the order of interfaces (across and within platforms) and tasks was counterbalanced. The second stage consisted of a “final survey”, which required participants to compare desktop and mobile interfaces.

Table 12  Phase 4 News Categories and News Topics

<table>
<thead>
<tr>
<th>Versions</th>
<th>Task Type</th>
<th>News Categories</th>
<th>News Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile</td>
<td>Practice</td>
<td>Sports</td>
<td>Check News about the 2018-19 NHL season</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>Check recent news about Colorado school shooting</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>Find news report thunderstorms bring flooding to Houston area</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Business and Finance</td>
<td>Find news about Uber IPO</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Business and Finance</td>
<td>Stocks plunge after Trump threatens new tariffs on China</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Health</td>
<td>Asthma sufferers failing to take life-saving medication properly</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Health</td>
<td>Find news about recent Ebola virus outbreaks</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Science and Technology</td>
<td>Find news about latest Google Pixel 3A</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Science and Technology</td>
<td>Stanford Studies Pokémon's Effect on Our Brains</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Sports</td>
<td>Check out the latest news from the UEFA Champions League</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Check News about the 2018-19 NBA season</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Find news about the Met Gala</td>
</tr>
<tr>
<td>Mobile</td>
<td>Real</td>
<td></td>
<td>Find news about the new Avengers movie</td>
</tr>
<tr>
<td>Desktop</td>
<td>Practice</td>
<td>Sports</td>
<td>Check recent world cup situation</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>Check recent severe weather damage and warning: Storm, Tornado</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Government, Politics, and Events</td>
<td>Find news report US North Korea Summit recent progress</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Business and Finance</td>
<td>You want to search why Two Melons Just Sold For $29,000 In Japan</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Business and Finance</td>
<td>Starbucks Is Closing All Its U.S. Stores for Diversity Training Day</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Health</td>
<td>The Truth of Swimming Pool and Hot Tub Germs</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Health</td>
<td>Which is healthier: hot dogs or hamburgers</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Science and Technology</td>
<td>Find news about the first all-screen smart phone (Lenovo Z5)</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Science and Technology</td>
<td>Microsoft is acquiring GitHub</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Sports</td>
<td>You want to check National Collegiate Athletic Association College baseball (NCAA) situation</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Sports</td>
<td>Check News about the New York Yankees</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Find news about the new Star Wars movie; Episode IX</td>
</tr>
<tr>
<td>Desktop</td>
<td>Real</td>
<td>Entertainment, Art, and Culture</td>
<td>Check news about 'Harry Potter' Star Matthew Lewis (Neville) Confirms Marriage to Angela Jones</td>
</tr>
</tbody>
</table>
The procedure for the study stage was the same regardless of platform. Specifically, there were four main steps each participant followed in the experiment. These four steps were i) completing an initial questionnaire, ii) performing tasks, iii) completing mid-study questionnaires, and iv) completing a post-study questionnaire.

The first step was a demographic questionnaire that included self-reported language proficiency in terms of reading, listening, and writing ability. Proficiency was measured on a scale from 1 to 4, with 1 corresponding to very limited proficiency, and 4 corresponding to native proficiency.

Next, users interacted with the previously described interfaces. More specifically, for each interface, users performed one practice task, followed by 3 tasks. For each task, participants were asked to issue their own queries to look for information on the given news topic, with the topic being displayed in all of the participant’s languages. Participants were also asked to mark any results they found relevant to the news topic content, using buttons placed next to each result (see Fig. 1-4). These markings were recorded by the experimental platform, along with task times, number of queries, as well as number of viewed documents (i.e. news results clicked/opened). Participants were free to move on to the next task if they felt that they had received sufficient relevant information on the task. Upon completion of all 3 tasks with one interface, a mid-study questionnaire was presented, which asked users to indicate their perceived level of search support provided by the interface.

Lastly, after having interacted with all interfaces (3 interfaces in desktop version and 2 interfaces in mobile version), a post-study questionnaire asked participants to indicate their overall comparative impressions, including i) which interface was ‘easiest/easier to use’, ii) which interface they ‘preferred the most/more’, and iii) which interface they ‘disliked the most/more’. In addition, participants were asked to indicate their reasoning for particular preferences/dislikes.

6.2.5 Data Analysis

All results were analyzed using Chi-squared tests and ANOVA tests (as well as Bonferroni-corrected posthoc tests), and each test used an alpha level of 0.05.
6.3 Results

6.3.1 Participant Demographics

For this study, 40 qualified polyglot participants were recruited, i.e. participants who are proficient in at least two languages. Half of the participants was proficient in Chinese and English, while the other half was proficient in Spanish and English. The participants had an age range from 14 to 58. The figure below shows the count of participants across different languages based on participants’ first language (L1) and second language (L2). The most common first language was English (16), followed by Simplified-Chinese (16), Spanish (7), and Traditional-Chinese (1). The most common second language was English (24), followed by Spanish (13), and Simplified-Chinese (3). All participants indicated native proficiency in their first language, and the majority of participants indicated high proficiencies (level 3 or 4) in their second language across reading, writing, and listening skills (39/35/37 respectively).

![Figure 46 Participants L1 and L2](image)

6.3.2 Results from Desktop Interface Comparison

6.3.2.1 User Preferences

As previously mentioned, user preferences were gathered through a comparative post-study questionnaire (i.e. after all tasks with all interfaces had been completed), as well as mid-study questionnaires (i.e. after all tasks for a particular interface had been completed).

In the post-study questionnaire, participants generally indicated clear preferences or dislikes towards specific interfaces, with the interface type having a statistically significant effect on ‘most preferred’ and ‘easiest to use’ answers. As shown in Figure 47, the ‘Panels’
interface was strongly considered to be the ‘easiest to use’ compared to all other interfaces (p-value<0.00001), and the comparatively ‘most preferred’ (p-value<0.00001) interface. The ‘most disliked’ interface was the ‘Tabbed’ interface (p-value<0.06).

![Bar chart showing user preferences for different interfaces.]

**Figure 47** Post-study questionnaire results. Desktop ‘Panels’ is considered the ‘most preferred’ and ‘easiest to use’ by most participants, while desktop ‘Tabbed’ is considered the ‘most disliked’.

In addition to these quantitative user preference analyses, a thorough investigation of participants’ qualitative answers was performed. When comparing all three desktop interfaces, ‘Panels’ was considered the easiest to use because participants felt they could better see search result lists of two languages displayed side by side at the same time, with a clear language indication and without requiring much effort (e.g. having to scroll down/toggle/click to find the results in the other language). It was found to allow participants to “easily and quickly compare” search results of both languages, for example, to find which article is interesting/relevant/unduplicated, or identify if the search is successful or the participant should try a different query. Moreover, it allows a participant to move to additional search results within a language without affecting another language. Overall, participants thought it is more “fit for a wide screen”. The ‘Universal’ interface was selected by some of the participants as the most preferred because it was found to have a clear layout for both languages on the same page, with the most recent and relevant news to compare without distraction, and without a “cramped layout” compared to ‘Panels’. ‘Tabbed’ was sometimes selected as the easiest to use because some participants indicated that they could
easily use the method of toggling between languages or select their preferred language or article on the desktop with control and without being visually distracted.

In terms of reasons for selecting their ‘most preferred’ interface, some participants thought that the ‘Panels’ interface is better due to its “easiness” and “efficiency”, which are similar to the reasons described in the previous paragraph relating to reasons for choosing the 'easiest to use' interface. Also, participants indicated that ‘Panels’ is compact with lots of information, and both the keyword and the translated keywords are clearly indicated. One participant even strongly expressed the preference "It is a unique interface with the potential to facilitate search for multilingual users. I would choose this over Google's interface.” The ‘Tabbed’ interface was liked by some participants because the tab toggle design was considered easy to use, allowing participants to quickly choose/switch between languages. It was also found that the toggle option to decide which language to choose/hide enables the participant to see results of one language at a time and easily find the information they want or go deeper into a specific language. Participants who preferred ‘Universal’ complimented on its comfort and aesthetics, as well as the fact that it is similar to the interfaces they typically use (e.g. Google, Bing, etc.). In addition, it does not require them to change back and forth between different languages. It was also considered to have the right balance of the two languages, especially if the user is significantly stronger in their first language than their secondary language. It has an advantage over traditional interfaces because it presents all the news in multiple languages at the same time but with a sequence, which makes the reading and comparison easier. For example, participants can read several news articles in English first and have a brief idea about the news they searched and understand the main content, and then look for relevant news in another language easily afterwards via scrolling down.

As previously mentioned, ‘Tabbed’ was considered the most disliked interface overall, followed by ‘Universal’ and ‘Panels’. A key reason mentioned for ‘Tabbed’ was the design of swapping between languages as it was considered inefficient, confusing, and distracting. It “requires extra time” from the user to switch between tabs and wait for the results to load, which would confuse first-time users and also “create distractions” since it does not display results of both languages at once (also making comparisons more difficult). It may also cause a user to forget to check news in other languages when the user thinks s/he had already gathered enough information in one tab (thereby potentially leading to incomplete information).
When ‘Universal’ was voted as most disliked, it was mentioned that it displays insufficient amount of results per language on one screen, and that it needs extra effort to scroll down to see more results (especially since most users only focus on the first few results when searching, and thus most of the time users may not notice the other language’s search result at the second half of the page). Moreover, it was found to be hard to compare the results of two languages, and it suddenly switches to another language midway, which was found to break the reading flow.

The ‘Panels’ interface was generally considered well-organized, but with some disadvantages as some participants shared that this type of interface is too “narrow/cramped”, and that it breaks the reading flow, which makes it not comfortable to read on a desktop. One participant highlighted that "When I search news, most of the time I just quickly briefly glance at the topic titles of the search result lists and click the most relevant one. This panel version interface breaks down the sentences and topics into more than one line, which makes it uncomfortable for quick reading." Some other participants indicated that the information provided by this interface is overwhelming and distracting when reading articles side by side in both languages. Some of them thought that since the two sides of ‘Panels’ actually display different content (i.e. not just direct translations of each other), which is hard to compare.

For the mid-study questionnaires (i.e. after having finished all tasks with one particular interface), results similarly indicated differences between interfaces, although less pronounced, as shown in Figure 48. In particular, differences between interfaces were found to be statistically significant for only some of the questions, including Q3 (‘helping combine information’) and Q5 (‘getting an overview’). Moreover, through Bonferroni-corrected post-hoc tests, it was found that both ‘Panels’ and ‘Universal’ interfaces were considered to be particularly good at ‘helping combine information’ compared to the ‘Tabbed’ interface (p-value<0.003 and p-value<0.009 respectively). Regarding ‘getting an overview’, ‘Panels’ had higher ratings compared to the ‘Tabbed’ interfaces (p=0.03).

Additional analyses with respect to participants’ second language proficiency levels did not yield any statistically significant results, suggesting that the preferences were not influenced by participants’ language skills in their second language. However, as previously mentioned, language proficiencies of participants were not well distributed among the participants of this user study, as the vast majority of participants indicated a proficiency level of 3 or higher in their second language, and thus further research needs to be conducted.
with a set of participants with second language proficiencies well distributed across lower proficiency, mid-level proficiency, and higher proficiency.

**Q1:** “The system provided **enough information** to help me solve the search tasks.” (p-value<0.12)

**Q2:** “The system provided me with **many different kinds of information**.”
(p-value<0.73)

**Q3:** “The presentation of search results helped me **easily combine information** from multiple languages.” (p-value<0.001, p=2.252030e-05)

**Q4:** “The presentation of search results allowed me to **easily identify relevant information**.” (p-value<0.55)

**Q5:** “The presentation of search results helped me get **an overview of the information** available in multiple languages.” (p-value<0.001, p=7.945266e-03)

Figure 48 Mid-study questionnaire results. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree).

### 6.3.2.2 General User Behavior

Participant behavior differences in terms of average task time, number of queries, and number of documents viewed were not found to be statistically significant (p-value<0.6/p-value<0.3/p-value<0.7 respectively) across the three interfaces.

Interestingly, participants' language proficiency in their L2 had a statistically significant effect on the number of queries issued (p-value<0.05). In particular, participants with high
proficiency score (4) issued a higher number of queries compared to participants with a low proficiency score (2 or 3).

Figure 49  Participant behavior features per interface

Figure 50  Participant L2 proficiency and average query number
6.3.2.3 Influence of Interface and News Topics

A series of analyses was conducted to investigate the influence of news categories (i.e. “Business and Finance”, “Entertainment, Art, and Culture”, “Government, Politics, and Events”, “Health”, “Science and Technology” and “Sports”), as well as any interaction effects with interface type. A statistically significant effect was found for news categories on number of queries (p-value<0.03), with “Science and Technology” tasks generally taking more queries than all the other categories. In addition, there was a marginally statistically significant difference regarding the news categories with respect to task time (p-value<0.066), with “Science and Technology” tasks having a longer task time than other tasks. Interestingly, there was no statistically significant difference in the number of viewed documents (p-value<0.09) with respect to news categories. As shown Figure 51, participants read a higher number of documents during ‘Entertainment, Art, and Culture’ tasks.
Figure 51  Participant behavior features per news category
6.3.3 Results from Mobile Interface Comparision

6.3.3.1 User Preferences

Based on the responses in the post-study questionnaire, participants generally indicated clear preferences/dislikes towards specific interfaces, with the interface type having a statistically significant effect on ‘more preferred’ and ‘easier to use’ post-study questionnaire answers. As shown in Figure 52, the ‘CollapseAll’ interface was strongly considered to be ‘easier to use’ and ‘more preferred’ (p-value<0.004, p-value<0.0001 respectively) compared to the ‘Tabbed’ interface. The ‘more disliked’ interface was the ‘Tabbed’ interface (p-value<0.02).

![Figure 52](image)

**Figure 52** Post-study questionnaire results. ‘CollapseAll’ is considered ‘more preferred’ and ‘easier to use’ by most participants, while ‘Tabbed’ is considered ‘more disliked’.

Besides these quantitative user preference results, we also performed a thorough investigation of participants’ qualitative answers. When comparing these two interfaces on mobile, ‘CollapseAll’ was considered “easier to use” since it presents the search results of both languages at once on the same page in a “well-organized format” with a “clear language indication” which is “easy to read and compare results”. This was found to enable users to quickly decide if they need to change their search keywords, since they could quickly glance over results from both languages. Besides, the collapse button was praised, as it was found to help users focus on the results of one language via hiding unwanted languages. One participant replied, "It divides the news into different sections so that I can know which region/country it is coming from. Since some news is region sensitive, this interface will help me to recognize and pick the most relevant news quickly." Another participant mentioned, "It
was interesting to see the different search results for the different languages and how different the results were! I found that using the English language resulted in more search results. That way, when I searched something in English, the interface would yield results in Spanish, and I can modify if needed." One participant even mentioned, “it allowed me to see both languages and, again, I really liked the fact that you could hide the information in the universal interface but still have it handy.” The mobile ‘CollapseAll’ interface was indicated as a good fit for the mobile since it only required participants to touch/scroll down by fingers in a more natural way compared to toggling tabs, which helped participants avoid extra effort and news page loading/waiting time per language. However, ‘Tabbed’ was still selected as the easiest one by some participants, who indicated that they could easily use the method of toggling between languages to select their preferred language with control, and that the layout of this design fits better on the screen since it is cleaner without showing too much information.

Most participants also selected the ‘CollapseAll’ interface as “most preferred”, with reasons including its easy and fast navigation, getting all information on the same page, no extra loading time to check for another language’s results, clear design (i.e. to better choose a preferred language, select more relevant news, compare and get different points of view), and more natural operation on mobile device (i.e. scrolling method instead of tabbing). Some participants, however, favored the ‘Tabbed’ interface due to the interface’s better user experience resulting from its easy-to-use, well-organized, clean and natural layout with results of one language per page, clear language tab indication, and easy tab toggling.

The ‘Tabbed’ interface obtained the majority of votes for the ‘more disliked’ interface, since its interaction style was considered unnatural and hard to operate for mobile devices, and causing confusion for participants. Due to the tab’s size and position on the mobile device, the tab usually got ignored, and, when participants switched tabs, it needed additional loading time to display results. The ‘CollapseAll’ interface was disliked by some participants because it looked “cluttered” on the mobile screen, which made participants uncomfortable and confused to see the mixed languages’ results. Besides, scrolling is also an essential factor that participants dislike, given users’ unwillingness to scroll down enough for the secondary language search results, which potentially leads to missing information.

Mid-study questionnaires result also indicated differences between different mobile interfaces, although less pronounced as compared to the results about desktop interfaces, as
shown in Figure 53. As in the desktop study, differences between interfaces were found to be statistically significant for Q3 (‘helping combine information’) and Q5 (‘getting an overview’). Results showed that the ‘CollapseAll’ interface was considered to be particularly good at ‘helping combine information’ and ‘getting an overview’ compared to the ‘Tabbed’ interface (p-value<0.0001 and p-value<0.0001 respectively).

**Q1:** “The system provided **enough information** to help me solve the search tasks.” (p-value<0.4)

**Q2:** “The system provided me with **many different kinds of information.**” (p-value<0.07)

**Q3:** “The presentation of search results helped me **easily combine information** from multiple languages.” (p-value<0.0001)

**Q4:** “The presentation of search results allowed me to **easily identify relevant information.**” (p-value<0.16)

**Q5:** “The presentation of search results helped me get **an overview of the information** available in multiple languages.” (p-value<0.0001)

**Figure 53** Mid-study questionnaire results of mobile interfaces. Ratings indicate the level of agreement with the statement (1=strongly disagree, 5=strongly agree).
6.3.3.2 General User Behavior

Participant behavior differences in terms of average task time, number of queries, and number of documents viewed were not found to be statistically significant (p-value<0.9/p-value<0.7/p-value<0.25 respectively).

![Figure 54 Participant behavior features per mobile interface](image)

Likewise, participants’ second language proficiency (read/write/listen) did not have any significant impact on their behavior.

6.3.3.3 Influence of Interface and News Topics

Regarding the influence of news categories (i.e. “Business and Finance”, “Entertainment, Art, and Culture”, “Government, Politics, and Events”, “Health”, “Science and Technology” and “Sports”), a marginal statistically significant effect was found for news categories on task time (p-value<0.06), with “Health” category tasks generally taking more time than all the other news categories. There was no statistically significant difference in the number of viewed documents (p-value<0.15) with respect to news categories or the number of queries (p-value<0.17) with respect to news categories. As shown in Figure 55, participants tended to issue slightly more queries while conducting “Health” and “Entertainment, Art, and Culture” tasks, and they read a higher number of documents during “Science and Technology” tasks.
Figure 55  Participant behavior features per news category
6.3.4 Comparison Between Desktop and Mobile Interfaces

After participants finished all tasks with the desktop and mobile interfaces, an additional overall survey was conducted to gauge comparative evaluations. When asked which platforms they preferred overall to check news (p-value<0.0001), most participants chose ‘Desktop’ (81%), followed by ‘Both’ (16.7%) and ‘Mobile’ (2.4%).

For the participants who chose ‘Desktop’, they responded that these interfaces enabled easier news reading, easier button clicking, typing, and more responsive interaction due to the desktop’s larger screen and faster response time. A number of representative excerpts of participants’ responses are recorded in Excerpts 1 and 2 below.

“The buttons on the mobile platform is not as big as I preferred, desktop is easier to type, and the bigger screen show more information.” -Excerpt 1.

“The Mobile version is not suitable for too much information or too many colors within one page. It makes the word size become smaller, which is hard to recognize. Further, the high complexity and multi-function cause distraction and loss of focus. On the other hand, the display on Desktop is much clearer. I tend to click different buttons more when use Desktop.” -Excerpt 2.

The remaining participants shared different specific reasons, including interface layouts, news search result relevance, personal habits, and robust functioning of the elements on the interface. In particular, the layout of the interface was regarded as an important factor in user’s consideration when choosing between desktop and mobile interfaces. Four participants thought the layout on mobile interfaces was not well suited for mobile platforms. Two participants in particular noted that the layout of the mobile interface did not carefully enough consider the small size of the screen, and thus made it hard to read and click. Likewise, one participant responded that the layout is the same for both desktop and mobile, and that this layout is more suitable for desktop.

User habits also played an important role, as they have developed a habit about what device to use for different tasks, as one participant responded (see excerpt 3).

“I don't typically use search engines on mobile as much as on desktop. Mobile is convenient as well, but I would prefer on desktop”. -Excerpt 3.
Regarding interfaces that were particularly liked, two participants specifically called out that they liked the ‘Panels’ interface compared with other interfaces on desktop (see Excerpt 4). For mobile, one participant specifically mentioned that the ‘CollapseAll’ interface is the easiest to use.

“I would like to see the Panels interface used more in general. I find it convenient and easier to use than other popular interfaces on desktops.” - Excerpt 4

Among the seven participants who chose both desktop and mobile, one responded ‘Both’ are “easy to use”, four responded that it depends on which platform is convenient for them at the time period they need to search for news (see excerpts 5, 6), and that they did not explicitly relate this to the interfaces of desktop or mobile. The remaining two participants chose ‘Both’ because they use mobile more for this type of searches although they mentioned desktop was “a little easier to use”.

“I like them both. which platform I prefer to use depends on which instrument, my laptop or mobile phone, I am more convenient to use for searching.” - Excerpt 5.

“I use them both very often, when I am in front of my computer, I normally use desktop, but if I am not, I use my phone which is always handy.” - Excerpt 6.

One participant liked the ‘Mobile’ because the participant thought it was quicker to navigate.

![Figure 56 Participant platform preferences](image_url)
In addition to overall preferences, we also enquired about specific interface preferences with respect to different platforms (i.e. desktop ‘Tabbed’ vs. mobile ‘Tabbed’, and desktop ‘Universal’ and mobile ‘CollapseAll’). When comparing the ‘Tabbed’ interface between desktop and mobile (p-value<0.0001), the desktop version was generally considered the more preferred version (83.3%) as opposed to the mobile version (16.7%). Popular reasons for choosing Desktop ‘Tabbed’ include ‘Tabbed’ being more “intuitive” on desktop, “clearer” and “easier to read”, “bigger” and “easier to click”, and “capable of displaying more information on screen” (see excerpt 7).

“When using mobile device, people like simply rolling down, instead of click. When using desktop, buttons are easily recognized and preferable to click. Therefore, the switch between two languages will be more reliable when on Desktop.” - Excerpt 7

In addition, ‘Tabbed’ was considered to be more suitable for desktop use cases, as users mentioned that they tend to scroll more on mobile given the smaller screen size.

Various reasons caused some participants to choose mobile ‘Tabbed’, most of which were related to each individual’s usage preference. Some participants liked the compact view and easiness of scrolling fast clicking. One participant responded, “the mobile version is compact yet clear”, and another participant mentioned “The mobile was easier to switch back and forth than the desktop”.

![Figure 57 Participant interface preferences: Desktop ‘Tabbed’ and Mobile ‘Tabbed’](image)

Figure 57 Participant interface preferences: Desktop ‘Tabbed’ and Mobile ‘Tabbed’

When comparing the desktop ‘Universal’ interface and mobile ‘CollapseAll’ interfaces (p-value<0.07), the mobile ‘CollapseAll’ (64.3%) was found to be preferred over the desktop version (35.7%). Some of the participants found that the mobile ‘CollapseAll’ makes the
divide between search results of different languages clearer, and having the option to collapse the search results of one language enables users to focus on the results of a certain language or read and compare the results of both languages at the same time. In addition, mobile ‘CollapseAll’ was said to allow users to scroll less during the reading, which reduces users’ efforts. Other participants who preferred desktop ‘Universal’ responded that they liked reading results of both languages without the need for collapsing, and that it is easier to go to next pages and close a website once opened on Desktop ‘Universal’.

Figure 58  Participant interface preferences: Desktop ‘Universal’ and Mobile ‘CollapseAll’

Recall that in the desktop study, there were actually three interfaces, with ‘Panels’ being the most preferred interface. However, when only comparing the desktop ‘Universal’ interface and desktop ‘Tabbed’ interface (p-value<0.8), the desktop ‘Tabbed’ (52.7 %) had slightly more votes than desktop ‘Universal’ (47.6%). Participants preferred desktop ‘Tabbed’ because it was found to be more “intuitive”, “flexible”, and “easy for switching” instead of having to scroll down, “clear” and having a “better organization” of the results from different languages instead of the “cluttered layout”. In addition, the desktop ‘Tabbed’ was found to have larger buttons, and users could quickly scan top search results of different languages by clicking a button, because “the top results are usually the ones users would like to read”. Popular reasons for choosing desktop ‘Universal’ included the preference for scrolling over switching tabs to see results of different languages, the capacity of displaying results of different languages at the same time after an initial search (“so it is easier to make the comparison”) and enable users to get an overview in a short time.
Figure 59 Participant interface preferences: Desktop ‘Tabbed’ and Desktop ‘Universal’

Comparing all five interfaces across both desktop and mobile systems (p-value<0.0001), the most popular interface was desktop ‘Panels’ (61.9%), followed by desktop ‘Universal’ (11.9%), desktop ‘Tabbed’ (21.4%), mobile ‘CollapseAll’ (4.8%), and mobile ‘Tabbed’ (0%). The majority of participants who preferred desktop ‘Panels’ credited its “clear”, “organized”, and “easy to compare” interface, which also “best utilizes the big screen nature of desktop”. A second common reason was related to the actions from users, as they liked the straightforward interaction of “search-once and see results without clicking or scrolling”. One participant also suggested that this could be applied to compare articles and see what language the user would like to read. Desktop ‘Tabbed’ was commended mostly on its “concise”, “easy to use”, and “visually pleasing” interface, which was found to be designed in a way that enables users to easily see results of a specific language they wanted. Participants mentioned that the ‘Tabbed’ design was suitable for desktop, and that it achieves a good balance between the two languages with the default being search results of the first language, and the results of the second language being in the next tab. Desktop ‘Universal’ was commended by some participants on its provision of information from different languages via simple scrolling. One participant responded, “Easy use, least click, simple rolling, but still present enough information”. Participants who preferred the mobile ‘CollapseAll’ interface shared that its collapse feature presented “more organized information”, and that it is “easy to tell the change in the language since it is bolded.”
Lastly, there appeared to be a common trend (with statistical significance) towards longer task times, higher number of queries, and higher number of viewed documents for the desktop versions compared to mobile versions (p-value<0.001/p-value<0.001/p-value<0.0001 respectively).

Figure 61  Participant behavior features: Desktop and Mobile
6.4 Summary and Discussion

This section includes a summary and discussion of the key findings from the phase 4 study on user preferences and interaction behaviors regarding desktop and mobile news search engine interfaces (RQ4.1 and RQ4.2), as well as the effects of news topics on interaction behaviors (RQ4.3). We will compare these results with prior work, provide further insights and recommendations based on participants’ free text responses (given as part of the post-study questionnaire), and discuss the implications for designing multilingual search interfaces.

In terms of platforms, the desktop versions were preferred significantly more, even though they lead to longer average user task times, a greater number of queries, and a higher number of viewed documents compared to the mobile versions. Participants mainly claimed two reasons. Firstly, participants commended the desktop versions’ organization enabling easier news reading, easier button clicking, typing, and more responsive interactions due to the desktop’s much clearer larger screen and faster response time. Secondly, participants mentioned their own habits, with desktop often being the preferred platform to perform such tasks.

When comparing specific interfaces across platforms, the desktop ‘Panels’ interface was generally considered the easiest to use due to its side-by-side layout providing clear language indication, requiring minimum searching/reviewing effort, presenting well-organized results, and enabling easy multilingual news results comparisons. These results are consistent with findings in phase 1. However, some participants still mentioned some disadvantages of this ‘Panels’ interface. In particular, several users claimed it was not comfortable to read since the interface was too narrow/cramped, producing information overload, as well as causing distractions and breaking the reading flow.

The desktop ‘Universal’ interface was considered the least preferred and least easy to use interface overall, even though it has clearly separated language sections like the desktop ‘Panels’ interface (with the only difference being that ‘Universal’ only has vertical sections, whereas ‘Panels’ has both vertical and horizontal blocks). However, some users considered it to be less cluttered compared to ‘Panels’, and making easier to get a good overview of information in a short time, particularly compared to ‘Tabbed’. This may explain why it was not as frequently mentioned as a disliked interface, especially compared to ‘Tabbed’, which is also consistent with the results found in phase 1 [5]. However, some participants
complained that there are not enough articles per language section (even though this was the same across all interfaces), and that requirement to scroll down cost too much extra effort and time. This was found to sometimes lead participants to not notice the other language result section at the bottom of the page.

The desktop ‘Tabbed’ interface was the most controversial interface, as it was voted second for ‘easiest’ and ‘preferred’ interface, yet first for ‘most disliked’. Participants who preferred desktop ‘Tabbed’ commended the ability to control the ease of reading, as they could fully concentrate on news in one language without being visually distracted. However, other participants mentioned that the tab button’s was inefficient (e.g. requiring extra time to switch between tabs and waiting for the results to load), and that the interface overall was confusing and distracting (e.g. creating distraction since it does not display both language results at once, which means users cannot conduct comparisons). It was also mentioned that this interface may cause users to forget to check the news in another language, especially when they think that they have already gathered enough information in one tab.

When comparing desktop ‘Tabbed’ and desktop ‘Universal’, the most significant point of contention was related to switching method easiness, where some participants preferred the tab button of ‘Tabbed’ (since it is large enough to be noticed and can help them control the flow), while others liked the design of ‘Universal’, which required scrolling down rather than clicking on tabs.

When focusing specifically on mobile interfaces, the mobile ‘CollapseAll’ was considered the more preferred and easier to use interface when. Overall, it was considered a well-organized design with clear language indications, letting participants get an overview of the different results. Two particular reasons were mentioned for this preference, with participants finding it a better fit for mobile devices overall. The first reason was that it only requires participants to touch/scroll down using fingers, which was considered a more natural way compared to tabs, thereby saving effort and page loading time. The second reason was the function that allowed participants to hide and reopen a particular language, which was considered particularly convenient for the small screen size. However, some concerns were also mentioned by several participants, such as its layout looking cluttered on the mobile screen, and participants potentially missing the other language results at the bottom if they disliked the scrolling down operation.
By contrast, Mobile ‘Tabbed’ was the more disliked interface on mobile devices, and even receiving zero votes in terms of the most preferred interface when comparing all interfaces across both desktop and mobile. In participants’ free-text comments, it was noted that the tabbed language switching function was especially unnatural and hard to operate on a small phone screen, as it required too much extra effort. Some participants did prefer this interface, however, due to it providing a cleaner structure with a more traditional one-language-per-page layout, without showing overwhelming amounts information.

When comparing similar designs on desktop and mobile, we found that the ‘Tabbed’ design was more suitable for desktop, and that the ‘Universal’/‘CollapseAll’ design has better performance on mobile devices. For the pair of desktop ‘Tabbed’ and mobile ‘Tabbed’, both share the same tab language switching faction. However, the design is considered to be more suitable for desktop use cases, as users generally tend to scroll on mobile devices due to the smaller screen. In terms of the comparison between desktop ‘Universal’ and mobile ‘CollapseAll’, this design was preferred more on mobile, since its scrolling down function was considered more natural, and because this was considered to save effort when compared to the desktop version.

In terms of general user interaction behaviors, there were several interesting findings across all devices and interfaces. First of all, within the desktop version and mobile version comparisons, the most/more preferred interfaces had the highest/higher number of viewed documents. Specifically, when comparing the desktop interfaces, the ‘Panels’ was considered the more preferred by participants, yet ‘Panels’ had a high task time and high number of viewed documents compared to the others. Likewise, for mobile interfaces, ‘CollapseAll’ was considered the more preferred interface compared to ‘Tabbed’, yet ‘CollapseAll’ use led to slightly higher task completion times. This suggests that participants did not find this extra time to be wasted effort, but rather a welcome facilitation of further news exploration.

When analyzing the effects of news categories, we found one significant result (i.e., desktop number of queries, p-value<0.01), as well as several marginally significant results (i.e. desktop task time – p-value<0.07, desktop number of viewed documents – p-value<0.09, mobile task time – p-value<0.06). In the desktop versions, “Entertainment, Art, and Culture” and “Science and Technology” news categories had the longest task times, as well as the highest number of queries and viewed documents. In terms of mobile, “Health” had the
longest task time, “Entertainment, Art, and Culture” and “Health” had the highest number of queries, and “Science and Technology” had the highest number of viewed documents.

In terms of suggestions on how to improve our current multilingual news search designs, a total of 30 out of 40 participants shared several thoughts and suggestions (described in the coming paragraphs).

General suggestions about improving the systems included UI changes, such as providing different reading modes (e.g. dark mode for the interface). Specific feedback about mobile interfaces included making fonts and buttons bigger, and ensuring the mobile webpage was adaptable to the mobile resolution, particularly considering diverse screen size of mobile phones and also the news content it is displaying. Some participants mentioned that the closing button on the mobile versions was too small to click using a finger, and that mobile screen displays were not displaying the content in a well-readable format. Specific comments about desktop interfaces included adding a closing/collapsing tab or button for the desktop ‘Panels’ interface. Additionally, there were a number of interesting comments regarding the relation between the level of proficiency in language(s), and the propensity of being comfortable in using ‘Tabbed’ or ‘Universal’. For example, one participant shared that “I am much more comfortable reading English than Chinese so ‘Tabbed’ interface is good for me as I can have English as the default tab, but I think that if someone was close in level to both languages then they may prefer the universal interface”.

Suggestions about aspects besides interface design issues can be divided into i) content of the search engine results, ii) configurable news sources, iii) user study setup, and iv) user study prototype.

Eleven participants provided some feedback related to the content of the search engine results, including improving language translation accuracy, improving accuracy of relevant news in the search engine results, considering slang and cultural differences such as incorporating ‘Spanglish’ into search query for better search results, and increasing the number of search results for non-English languages.

In addition, there was interesting feedback regarding the need for more news sources or configurable news source preferences. For example, one participant responded that he “prefers Times and CNN instead of Yahoo news”.

---

109
There were also several suggestions regarding the user study setup. One participant suggested that study could have a mobile phone with a bigger screen size, as this would have better matched his daily habits. In addition, one participant suggested that it would help if the study were to eliminate the influence of search results’ contents on users, so as to focus on evaluating the best interface. This would require the user study to use the exact same topics and results for each of the different interfaces and platforms (which may, however, lead to learning effects).

Lastly, a number of comments were related to the user study experiment prototype, with one participant pointing out that there was a minor functional bug with the mobile interface, leading to some search results in the tabbed interface needing to be loaded again (and hence causing a delay). Likewise, some participants mentioned that the search engine speed is a bit slow compared to commercial search engines. This may be eliminated in future user studies by developing a more robust and responsive prototype.
Chapter 7

Conclusion and Future Work

With the global expansion of the Internet and the World-Wide Web, users have become increasingly diverse, especially in terms of language skills. In particular, there are now many users who are proficient in more than one language. However, such users often still face unbalanced and fragmented news access, even though they are able to read news from diverse information sources. This is in part due to the fact that most news systems are built under the assumption that users only consume news in one language. Moreover, there has been a distinct lack of research on how to best support multiple language abilities, particularly in terms of multilingual news search result pages. This thesis has provided strong evidence that multilingual users would appreciate an integrated news search engine platform, which can aggregate and display different language news content based on their preferences and language abilities, to help them gain access to more balanced news content.

In this thesis, I conducted research on multilingual news access through four phases. The first phase has evaluated four general Web search interfaces designed specifically with multilingual users in mind. Overall, it was found that clear language separation in a ‘Panels’ interface was strongly preferred, whereas an ‘Interleaved’ approach was least preferred. The findings from this interactive study thereby extend similar results found in studies that used static interfaces, and further present a compelling case for changing how multilingual search interfaces are built. In particular, while prior research in Multilingual Information Access has traditionally been confined to ‘Interleaved’ or ‘Tabbed’ presentation approaches, the study in phase 1.1 has shown that these approaches are in fact often the least preferred by multilingual users. In terms of interaction behaviors, it was found that users make significant use of each of their languages during search sessions. In particular, participants view and mark documents as relevant to a similar degree in each of their languages.

To study novel solutions for multilingual news access, the second phase of this thesis has presented the first large-scale investigation of multilingual news access behaviors. This investigation included two stages of user studies, which sought to survey current news consumption behaviors (stage 2.1), as well as news query and result selection behaviors (stage 2.2). The overall conclusions from stage 2.1 are that users currently already use
multiple languages when looking for news online, and that they do so particularly when using digital platforms. It was also found that different languages are used for different purposes (in terms of news topic domain), and that Non-English first language (L1) participants in particular make significantly more use of their second language (L2) language (which is most often English). However, it was also found that this multilingual behavior is currently not well supported by current available news access systems, and that significant burden is placed on the user to seek balanced and diverse news information. The main conclusions from stage 2.2 are that multilingual users make significantly more use of each of their different languages when a system provides multilingual support, such as when the system does not rely on users having to write queries in non-L1 languages (as the user’s proficiency may prevent them from doing so), and especially when a news search result page aggregates news from different languages.

In phase 3, I designed 2 pairs of mobile multilingual news interfaces, and then evaluated these through a pilot user study. Overall, participants preferred interfaces that provide an easy content comparison (from different languages), clear language indication, and easy hand access operation. ‘CollapseAll’ (which provided more information upfront) was found to be preferred over ‘CollapsePartial’ (which provided only a quick overview for each language), and a ‘Tabbed’ interface was considered more comfortable to use and more preferred compared to an interface that included ‘sliding’ between languages. When comparing all four interfaces, it was shown that ‘Slide’ and ‘Tabbed’ were both preferred over the ‘CollapseAll’ and ‘CollapsePartial’ interface designs, which may indicate that a single-language-per-page layout is more favored for mobile platforms. However, ‘CollapseAll’ was considered the easiest to use. The ‘Slide’ design was the most controversial interface, as it was voted first for ‘most disliked’ and the ‘most preferred’ interface. This indicates that there may be a strong split between user preferences. Across all interfaces, ‘Slide’ and ‘Tabbed’ interfaces had the shortest task completion time; at the same time, they were both voted as the ‘most preferred’ interface. Likewise, ‘CollapseAll’ had shorter task time and was more preferred than ‘CollapsePartial’.

Phase 4 extended the mobile news search interface comparisons from phase 3, as well as the desktop-based general search interface comparisons in phase 1.1, through an in-depth investigation of polyglot user preferences and interaction behaviors for multilingual news search engine interfaces on both desktop and mobile platforms.
In terms of platform preferences, it was found that the ‘desktop’ interfaces were significantly preferred over the mobile interfaces. The main reasons for this preference included ease of use, as well as users’ traditional habits of news searching and reading.

The desktop ‘Panels’ interface was the most preferred interface on desktop, as well as overall, when comparing all interfaces. This result was the same as the one found in phase 1, i.e. for general multilingual desktop search interfaces. Again, it therefore appears that clear language separation is preferred. When comparing all interfaces, the most frequently mentioned reason for this preference was the enhanced overview and ability to compare results from different languages. Likewise, some participants preferred desktop ‘Panels’ and mobile ‘CollapseAll’ because these interface layouts provided well-organized results, enabling easy comparison of multilingual news search results. However, as with other studies, some participants still favored desktop ‘Tabbed’ and mobile ‘Tabbed’ because these designs gave them the ability to control the language selection, and because they helped them to better concentrate on the news content without breaking the reading flow. Desktop ‘Universal’ may be considered as a compromise between desktop ‘Panels’ (which has overview/comparison benefits) and the desktop ‘Tabbed’ interface (which has a clear language separation). However, since the secondary language result list was on the page bottom, sometimes participants were found to ignore these results. The study in phase 4 also compared similar designs across desktop and mobile. It was found that a 'Tabbed' design was more suitable for desktop, whereas ‘CollapseAll’ was more preferred for mobile. As mentioned by several participants, this preference for ‘CollapseAll’ was potentially due to users being used to scroll (and not minding to do so) on mobile devices, whereas this is not the case for desktop interfaces (hence the less preferred desktop ‘Universal’ design).

Based on the research studies, results, and analysis that have been conducted in this thesis, there are multiple potential directions for future research.

Further user studies can be conducted for multilingual news search interfaces with a different set of participants whose L2 language proficiency is more distributed across all levels of proficiency (high, medium, low), in order to investigate how different levels of L2 language proficiencies influence users’ preferences and interaction behaviors on multilingual news search interfaces, since the user studies presented in this thesis had a participant population with mostly high L2 language proficiencies. In addition, it would be very interesting to investigate whether, and to what extent, languages themselves have an effect on
how the multilingual news search interfaces should be designed, in order to produce a more complete overview of multilingual news search across different parts of the world’s population, for example whether the preferred interfaces by users with Chinese/English skills, or Spanish/English skills would be liked by users of less frequently spoken languages.

Another interesting topic is to study whether different sizes of mobile screens have an impact on users’ preferences and behaviors on multilingual news search interfaces, as the mobile studies in this thesis were conducted with only one size of mobile screen (this was also mentioned by multiple participants in the study). Further user studies can be conducted with mobile phones with various screen sizes, as to allow participants to use a screen size that is the same as the one used in their daily lives.

Another opportunity lies with regards to auto-translation functions (e.g. providing translations of returned results), and future work could also be conducted on improving query translation accuracy, especially considering slang and cultural differences (e.g. including 'Spanglish' as mentioned by participants in phase 4). This may allow the evaluation of any potential effects on users’ behaviors (e.g. potentially increasing L2 document views).

Lastly, research can be conducted to evaluate how personalized news can be incorporated into multilingual news search engines and their interfaces. This may include machine learning algorithms that predict and/or aggregate the most relevant multilingual news content (in terms of languages and news source), as well as personalized interfaces for a given news query from an individual user. This could potentially be based on each user’s individual preferences and language abilities.
Reference list


