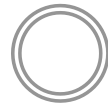


Babel revisited:
A taxonomy for ordinary images indexing in
a bilingual retrieval context



Elaine Ménard

&

Margaret Smithglass

School of Information Studies

McGill University

[Canada]

July 5th, 2011

Outline

2

- Introduction
- Objectives
- Methodology
- Results
- Discussion
- Conclusion

Introduction

3

Introduction

4

- Users do not necessarily seek the same image with the same concepts or the same terms.
- Systems using the low-level features are not yet widely available on the Web.
- Consequently, Internet users tend to mainly use text to formulate their queries.

Introduction

5

- Another level of difficulty surfaces when the language of the query is different from the language used for indexing.
- A user should be able to formulate a query to search for images in their native language, making the target language transparent.
- Detailed observations of the search behaviours are needed for a full understanding of the cognitive aspects involved in image searching.

Objectives

6

- Explore search behaviours of image searchers to learn about the terminology used and evaluate how this terminology can be incorporated into the development of a bilingual taxonomy for digital image indexing.
- Two research questions:
 1. In general, what types of terms and concepts are primarily used in queries formulated to retrieve images on the Web?
 2. What are the similarities and differences between the terms and concepts found in the queries made by the four linguistic communities?

Methodology

7

Sample

- 4 groups of 10 participants from 4 different linguistic communities
 - French
 - English
 - Russian
 - Chinese
- Participants did not have any professional experience in a field related to image indexing and retrieval.

Methodology

8

Data Collection - Phase 1

- Lab-based experiment with the 40 users
- 10 images were shown consecutively to all participants, who were asked to record all the appropriate queries they could think in order to retrieve each one of the 10 images.
- No limit to the number of queries was set.
- Each retrieval task was limited to five minutes per image.

Methodology

9

Data Collection - Phase 2

- Each participant completed a short questionnaire.
- 3 sections:
 1. Demographic questions
 2. Questions about the frequency of use of concepts used in image searches on the Web
 3. Questions about the frequency of use of terms in image retrieval searches on the Web

Sample

10

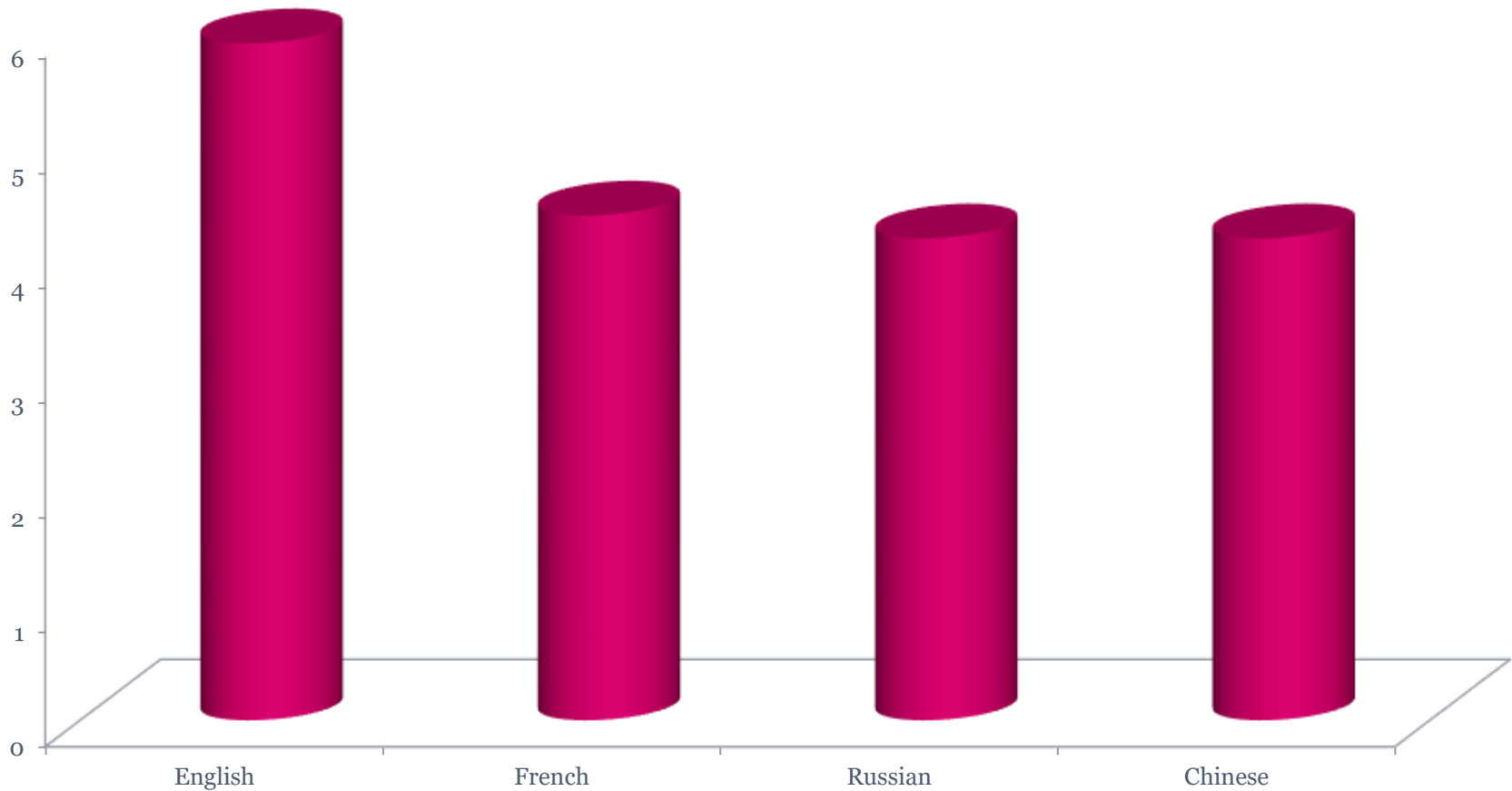
Participants

- Gender
 - 60.0% women
 - 36.0% men
 - 4.0% did not answer
- Age
 - 62.5% less than 26
 - 32.5% between 26 and 35
 - 5.0% between 36 and 45
- McGill University students
 - Variety of specializations and profiles
 - Various diplomas

Findings

11

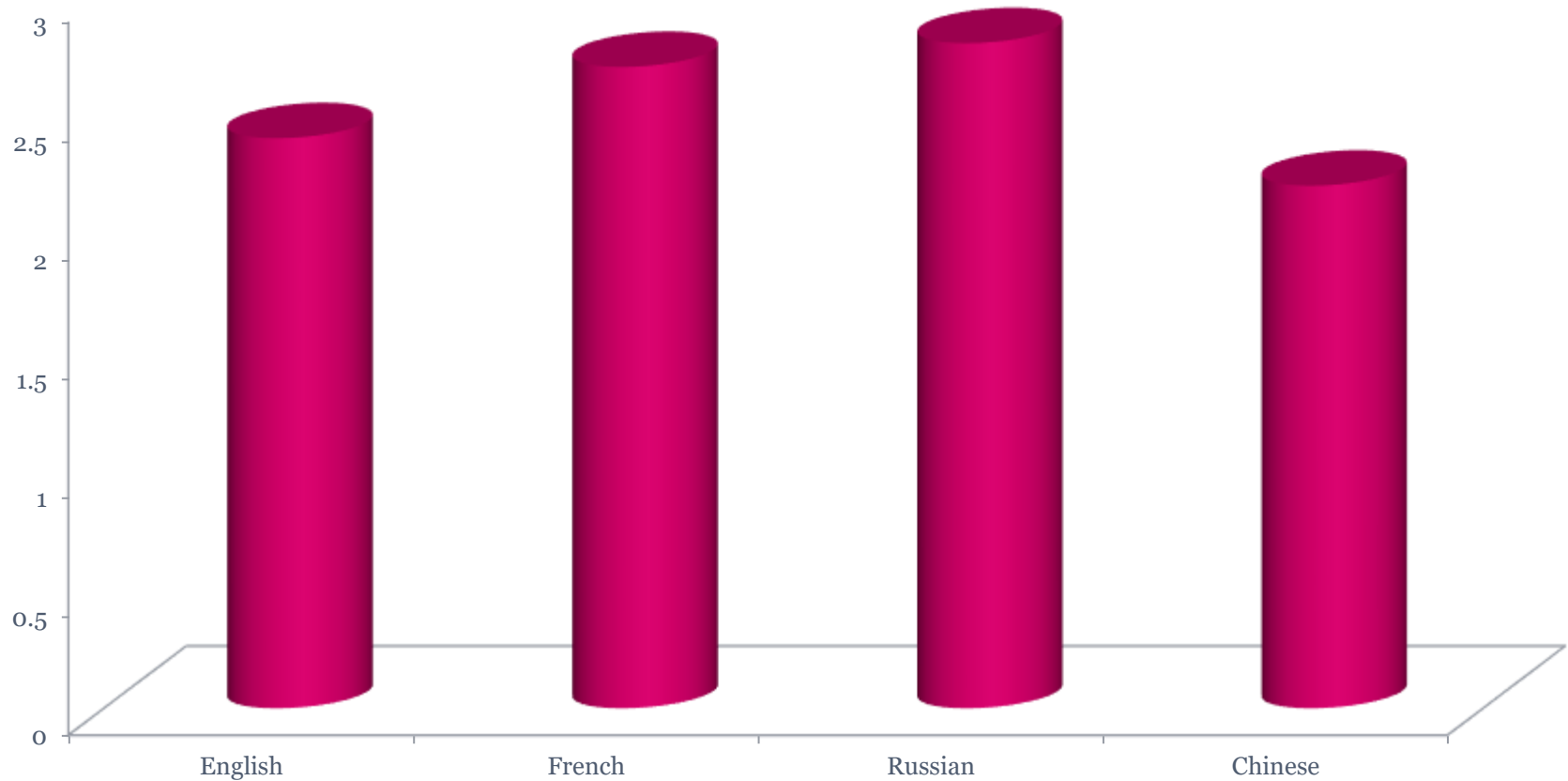
Average number of queries used per image



Findings

12

Average number of terms per query



Findings

13

Concepts – All participants

Frequently or very frequently used

- Trademarks
- Sponsors
- Person names

Rarely or very rarely used

- Emotions
- Atmospheres

Findings

14

Concepts – Comparison of the 4 groups

Conceptual Similarities

- Functions
- Activities / Actions
- Numbers / Dimensions

Conceptual Differences

- Colors
- Place names
- Events

Findings

15

Terms – All participants

Frequently or very frequently used

- Specific terms
- Compound names
- Color adjectives
- Proper names

Rarely or very rarely used

- Superlative adjectives
- Abbreviations
- Articles & adverbs

Findings

16

Terms – Comparison of the 4 groups

Terminological Similarities

- Compound Names
- Color Adjectives
- Dimensions

Terminological Differences

- Verbs
- Adjectives
- Proper Names

Findings

17

- English-speaking participants mainly confirmed that they rarely perform queries with terms extracted from another language.
- Other participants from the 3 other linguistic groups often formulate queries with terms extracted from a language different from their native language when performing image searches.

Discussion

18

- Difference in the length of the query according to the linguistic community.
- This is similar to previous studies on Web image queries.
- Participants from different countries, even if they are native speakers of one specific language, do not necessarily have the same cultural background and the same general knowledge.

Discussion

19

- Some concepts appeared regularly in the participants' queries.
- Similar to the results observed in our previous study (Ménard, 2008), which emphasized that image indexing should ideally be based on the pre-iconographic level.

Discussion

20

- Some concepts are only occasionally used or simply rarely used when searching for images.
- Even if emotional responses to image content can be considered as “universal”, image indexing with “emotions” and “atmosphere” remains a difficult process to achieve.

Conclusion

21

- The comparison with other populations speaking different languages could also enrich the knowledge on image indexing and retrieval performance in a multilingual context.
- The results of this research will contribute to the design of a new bilingual taxonomy.

Conclusion

22

- The new taxonomy is intended to be a powerful tool for use by cataloguers or indexers.
- The bilingual taxonomy will constitute a benefit for image searchers who are not very familiar with images indexed in English, which is still the dominant language of the Web.

Thank you!

23

elaine.menard@mcgill.ca

