

The Role of Language Skills in Interactive Social Book Search Sessions

Maria Gäde
Berlin School of Library and Information Science
Humboldt-Universität zu Berlin
Dorotheenstr. 26, 10117 Berlin
Germany
maria.gaede@ibi.hu-berlin.de

Mark M. Hall
Department of Computing
Edge Hill University
St Helens Road, L39 4QP
United Kingdom
mark.hall@edgehill.ac.uk

ABSTRACT

When searching for books, people frequently have to deal with content that is in a language different from their own. However, research on multilingual systems has generally focused on the user interface's language rather than the content language. In this paper, we describe and compare early results from the multilingual aspects in the Interactive Social Book Search (iSBS) task at CLEF 2014 and 2015. A preliminary analysis of usage patterns for native English and non-native English speakers indicates an influence of language skills on search behaviour during goal-oriented and casual leisure tasks. Based on previous experiences and results, strengths and challenges of IIR studies are discussed.

CCS Concepts

•Information systems → Users and interactive retrieval; Search interfaces; •Human-centered computing → HCI design and evaluation methods;

Keywords

Social Book Searching, Casual Leisure Behaviour, Multilingual Information Access, Interface Design

1. INTRODUCTION

Ideally, information systems provide boundless access to information, irrespective of the user's origin, linguistic background and search strategy [12]. While a lot of effort has gone into the implementation of multilingual user interfaces, less research has focused on the impact of the content language on the interaction between the user, the system, and the content [14]. Previous research assumed that differences in search behaviour are based on the level of cognitive effort the user needs to access a website [7]. It is thus likely that language skills should influence search strategies, and preferences [13].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '16, March 13-17, 2016, Carrboro, NC, USA

© 2016 ACM. ISBN 978-1-4503-3751-9/16/03...\$15.00

DOI: <http://dx.doi.org/10.1145/2854946.2854990>

Research in these areas has generally focused on professional work scenarios, where there is a specific, focused goal that the user is attempting to satisfy. However, particularly in book search contexts, users will frequently have no specific goal or only a very rough aim of what they might want to achieve in their search session and will frequently be searching in a personal and leisure context. To be able to support exploring and discovering strategies we need to understand the characteristics of such open-ended, leisure-focused sessions and possible language barriers within those.

2. INTERACTIVE SOCIAL BOOK SEARCH (ISBS)

The overall goal of the Social Book Search lab at CLEF (Conference and Labs of the Evaluation Forum) in 2014 [1] and 2015 [6] was to investigate how professional metadata (title, authors, ...) can be combined with social meta-data (tags, reviews) to satisfy an information need. Within this the Interactive Social Book Search Task (iSBS) looked at how the two types of meta-data can be combined in the search interface and in what way users make use of the two meta-data searches when interacting with the interface to complete a task.

To investigate this two search user interfaces (SUI) were created and two tasks defined that would be used with the two SUIs. The experiment was set up using SPIRES [5], which provides standardised pre- and post-task and pre- and post-session questions. The SUIs were created using PyIRE [4], which allows detailed data on the user interaction to be collected even if the experiment participants are located remotely.

Using the interfaces and tasks two user experiments were conducted using participants recruited from across Europe. In this paper we will focus only on the following research question:

RQ1 What role do language skills play in *goal-oriented* and *non-goal* tasks using an interactive multi-stage interface?

2.1 Tasks

In both experiments each participant used one of the two interfaces to complete both tasks. A latin-square setup was used to ensure a balanced distribution of interface use and task ordering. In both 2014 and 2015 participants had to complete a *goal-oriented* and a *non-goal* task. In both years the following *non-goal* task from [11] was used:

Imagine you are waiting to meet a friend in a coffee shop or pub or the airport or your office. While waiting, you come across this website and explore it looking for any book that you find interesting, or engaging or relevant.

In 2014 the following “simulated leisure task” [10] was used for the *goal-oriented* task:

Imagine you are looking for some interesting physics and mathematics books for a layperson. You have heard about the Feynman books but you have never really read anything in this area. You would also like to find an “interesting facts” sort of book on mathematics.

This generated very short sessions, thus to increase session length in 2015 the following *goal-oriented* task was used:

Imagine you participate in an experiment at a desert-island for one month. There will be no people, no TV, radio or other distraction. The only things you are allowed to take with you are 5 books. Please search for and add 5 books to your book-bag that you would want to read during your stay at the desert-island:

- Select one book about surviving on a desert island
- Select one book that will teach you something new
- Select one book about one of your personal hobbies or interests
- Select one book that is highly recommended by other users (based on user ratings and reviews)
- Select one book for fun
- Please add a note (in the book-bag) explaining why you selected each of the five books.

Additionally in 2015 participants were first given a training task that introduced them to the functionality available in the SUI they would be using to complete the two tasks.

2.2 Interfaces

Two SUIs were developed, with both providing a monolingual English interface [3].

The baseline (BL) interface implemented a standard SUI, consisting of a search box, the search result list, the item details display, and an area in which to collect the books that fulfill their task. It represents a standard interface that participants are likely to be familiar with.

The multi-stage (MS) interface implemented a novel SUI that consisted of three linked pages that implemented the pre-focus, focus, and post-focus phases of [12]. The first page provided an interface aimed at supporting less focused exploration of the collection, the second page a focused SUI that included the ability to restrict results by a given facet, and the third page a review interface that allowed participants to review the books they had collected and add or remove as required.

In 2015 the MS interface was updated and re-designed according to previous results and user comments [2].

Table 1: Number of participants by language using the multi-stage interface

Year	Native English	Non-Native (Other)
2014	8	14
2015	16	79

2.3 Data

For all tasks a monolingual English subset of the INEX Social Book Search’s Amazon/LibraryThing book collection was used, consisting of approximately 1.5 million books. Each book consisted of publisher-supplied meta-data (title, author(s), publisher, publication year, etc.), subject meta-data (classification codes, subject headings), user-generated content (Amazon user reviews, LibraryThing user tags), and a thumbnail image.

2.4 Participants and Data Gathering

In 2014 and 2015 participants from different countries and language backgrounds were recruited. A total of 233 users, 41 in 2014 and 192 in 2015, from 36 different countries and 30 different mother tongues participated in the experiment. Participants’ mother tongues included amongst others Afrikaans, German, Dutch, English, Danish, Romanian, Farsian, Russian, Turkish, Chinese or Portuguese. Those that did not select English as their mother tongue will for the purpose of the analysis be grouped together as non-native English speakers.

The SPIRE system’s latin-square functionality balances participants across the task/interface combinations, but this does not take into account the participants’ mother tongue. As a result in 2014 all but one of the English-language participants used the multi-stage interface, making a comparative analysis of the baseline SUI impossible. From this point forward we will only be considering the participants who used the multi-stage interface (Table 1).

The SPIRE system collected the following data points for each participant, which form the basis of the analysis:

- user profile (questionnaires), e.g. age, gender, level of education, first language, all languages used in web-search, country of residence;
- user – system interactions (from logs), e.g. queries, books collected, facets selected, UI elements interacted with;
- post-task motivation (questionnaires), e.g. why did you select these books, usefulness of UI elements, usefulness of meta-data elements;
- user engagement (questionnaires) based on [8].

3. THE IMPACT OF LANGUAGE SKILLS ON SEARCH BEHAVIOUR

To investigate the language impact, the log data collected during the participants’ use of the MS interface was automatically processed and the following five characteristics extracted for both of the tasks:

- Session length: Total length of the session;
- First interaction: Time from the interface loading to the first user activity;

- First query: Time from the interface loading to when the first query was executed;
- First book: Time from the interface loading to when the first book was collected;
- Books: Total number of books collected.

No tests were done for language impact on task ordering, search behaviour, or engagement, as [9] found no language impact on any of these aspects. However, in both experiments the qualitative responses of many of the German language participants stated that they encountered a significant language barrier when exploring the English language content, which is the motivation for the analysis presented here.

3.1 Goal-oriented Task

2014.

In the *goal-oriented* task English native speakers were faster in their first interaction with the system and faster to run their first query (Wilcoxon signed rank $p < 0.05$). No statistical differences were found in the time needed to collect their first book, the total number of books collected, and the total session length.

2015.

The only statistically significant difference was in the total session length (Wilcoxon signed rank $p < 0.05$). English native speakers' median session length was approximately four minutes faster than the non-native speakers'.

Comparing the two years, in the second year non-English speakers seem to have less difficulties dealing with the interface and task. The most likely reason for this is the introduction of the training task, which gives the non-native speakers the opportunity to get to know the SUI and reduces language and system knowledge impact on the initial interactions. This strongly indicates that for multi-lingual IIR experiments a training task is a necessity to minimise the language learning effects of the interface itself.

At the same time in 2015 the non-native speakers spent significantly more time on the *goal-oriented* task than the native speakers. Considering that there was no difference in the number of books selected and the initial speed of interaction, it is likely that the non-native speakers needed more time to inspect and select content. While not statistically significant, the 2014 experiment shows a similar trend, but due to the simpler task the individual differences mask the language impact. However, especially for casual leisure situations, longer session duration does not always indicate a poorer experience or dissatisfaction, but could also indicate a good user experience.

3.2 Non-goal Task

There are no statistically significant differences in the total session length in either year.

2014.

In the *non-goal* task English native participants were faster to issue their first query, collect their first book, and collected more books overall (Wilcoxon signed rank $p < 0.05$). However, they did not take less time for their first interaction.

2015.

Unlike in 2014, there is no difference on the time to the first query or collect the first book. However, English native participants were faster with their first interaction (4 seconds faster). Interestingly in 2015 the English native participants actually collected less books than the non-native participants.

As in the *goal-oriented* task, the introduction of the training task reduces much of the language impact of the interface. However, the slower time for the first interaction indicates that the impact of the content language remains. On the initial page, participants saw a hierarchical tree of topics that they could browse. While this had no impact in the *goal-oriented* task, in the *non-goal* task it seems that because participants have to come up with their own goal, the non-native speakers take a bit longer to translate their goal into the appropriate part of the topic tree, leading to a slower initial interaction.

The training instructions were designed so as not to have any direct connection to either the two tasks. To test this, we checked the queries issued by users in the training task against the queries used in the two main tasks. Only 6 of the 192 users (3%) reused at least one training query in either the *goal-oriented* (3 users) or *non-goal* (3 users) task. 5 re-users were English non-native speakers and 1 was an English native speaker. 3 participants re-used in their first task, 3 in their second. Overall this indicates that while the training instructions can act as direct prompts, potentially influencing the results, they can also simply act as reminders of topics the user is interested in in any case.

The differences in the number of books collected in the *non-goal* task is puzzling. The task did not change between the two years and none of the changes to the user-interface explain why in 2015 the non-native speakers collect more books than the native speakers, while in 2014 the situation is reversed.

4. CONCLUSIONS AND OUTLOOK

The results from two years iSBS data indicate differences between native and non-native speakers facing *goal-oriented* and *non-goal* casual leisure tasks. In particular they indicate that non-native English speakers require additional time to process the, for them, foreign-language content, which was to be expected.

However, the more interesting aspect is that the introduction of a training task in the second year reduced many of the language-specific differences. This gives a strong indication that it is possible to run experiments investigating multilingual issues in a mono-lingual environment, as long as participants are trained on the environment, so that only the content language impacts the results. There is of course a danger that the training task influences the results of the experiment itself, however our results show that only a small number of participants (3%) are influenced by the training task.

One of the difficulties with (multilingual) IIR studies is that they often struggle to reproduce and explain effects. In our analysis we see this with the number of books collected in the *non-goal* task. Here, although many non-native speaker participants reported finding the task difficult in both years, in the second year they collected significantly more books than the native speakers. A brief review of the annotations provided for most of the collected books also does not reveal

any issues such as collecting books they are not interested in just to complete the task.

One issue for further work that this raises is whether the metrics we are using are stable enough and have a clear enough interpretation to enable conclusions to be drawn. In the case of the books, perhaps in the *non-goal* task individual differences such as age or education are actually much stronger than the language aspects and the difference between the two years is pure chance. Other situation or context indicators might only appear during an experiment. Using thinking aloud protocols in smaller studies might help detecting and understanding the influence of additional factors we are currently not aware of.

Additionally as indicated [5] the variety of indicators and metrics that are used to characterise user behaviour remain a problem within, but also between studies. For example, session duration and selection of books could be interpreted in multiple ways. Both longer and shorter durations could be argued as being better. The same goes for the number of books. The user engagement scale [8] represents a stable metric across the two years of the experiment. However, [9] suggest that they see no language effects that could explain the differences we see in our results.

As far as the comparability between studies goes, the heterogeneity of the user groups, tasks, and interfaces, which create a more realistic setting and are one of the strengths of IIR compared to classic IR, also make comparison between studies difficult. Part of the iSBS track's goal was to create a common baseline and set of experiments that different research groups could use to investigate different research questions while maintaining comparability. However, as the results presented here show, we are still some way from creating such a stable baseline that would enable reliable comparisons. Therefore, the major challenge in IIR remains to identify and establish suitable interactive IR infrastructures that support comparable long-term studies.

5. REFERENCES

- [1] BELLOT, P., BOGERS, T., GEVA, S., HALL, M., HUURDEMAN, H., KAMPS, J., KAZAI, G., KOOLEN, M., MORICEAU, V., MOTHE, J., ET AL. Overview of inex 2014. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*. Springer, 2014, pp. 212–228.
- [2] GÄDE, M., HALL, M., HUURDEMAN, H., KAMPS, J., KOOLEN, M., SKOV, M., TOMS, E., AND WALSH, D. Overview of the sbs 2015 interactive track. In *CEUR Workshop Proceedings* (2015).
- [3] HALL, M. M., HUURDEMANN, H., SKOV, M., WALSH, D., ET AL. Overview of the inex 2014 interactive social book search track.
- [4] HALL, M. M., KATSARIS, S., AND TOMS, E. A pluggable interactive ir evaluation work-bench. In *European Workshop on Human-Computer Interaction and Information Retrieval* (2013), pp. 35–38.
- [5] HALL, M. M., AND TOMS, E. Building a common framework for iir evaluation. In *CLEF 2013 - Information Access Evaluation. Multilinguality, Multimodality, and Visualization* (2013), pp. 17–28.
- [6] KOOLEN, M., BOGERS, T., GÄDE, M., HALL, M., HUURDEMAN, H., KAMPS, J., SKOV, M., TOMS, E., AND WALSH, D. Overview of the clef 2015 social book search lab. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Springer, 2015, pp. 545–564.
- [7] KRALISCH, A. The impact of culture and language on the use of the internet: Empirical analysis of behaviour and attitudes, 2005.
- [8] O'BRIEN, H. L., AND TOMS, E. G. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.
- [9] SKOV, M., AND BOGERS, T. The influence of language proficiency on book search behaviour. In *CLEF2015 Working Notes* (2015).
- [10] SKOV, M., AND INGWERSEN, P. Exploring information seeking behaviour in a digital museum context. In *Proceedings of the second international symposium on Information interaction in context* (2008), ACM, pp. 110–115.
- [11] TOMS, E., AND HALL, M. The chic interactive task (chici) at clef2013. CLEF.
- [12] VAKKARI, P. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation* 57, 1 (2001), 44–60.
- [13] VASSILAKAKI, E., AND GAROUFALLOU, E. Multilingual digital libraries: A review of issues in system-centered and user-centered studies, information retrieval and user behavior. *The International Information & Library Review* 45, 1 (2013), 3–19.
- [14] ZHANG, J., AND LIN, S. Multiple language supports in search engines. *Online Information Review* 31, 4 (2007), 516–532.