

I Need More Time!: The Influence of Native Language on Search Behavior and Experience

Pengyi Zhang¹, Chang Liu¹, Preben Hansen²

1. Department of Information Management, Peking University
5 Yiheyuan Rd, Haidian District, Beijing 100871, China
{pengyi, imliuc}@pku.edu.cn

2. Department of Computer and Systems Sciences, Stockholm University
Borgarfjordsgatan 12, SE-164 07 Kista, Sweden
preben@dsv.su.se

Abstract. This paper describes our participation in the interactive track (ChiSwe Group) of the Social Book Search lab organized at CLEF 2016. This is our first participation in CLEF SBS interactive track. A total of 112 participants (29 native Chinese speakers 27 native English speakers, and 56 native speakers of other languages) participated in the SBS Interactive Track. We found that native Chinese speakers devoted more search efforts in searching, i.e. spent longest time to complete search tasks, selected the most number of books, have switched most between search, browse and review model of the search system, viewed more items and more metadata, and annotated more than native English speakers and native speakers of other languages. However, when evaluating the search engagements, Chinese speakers had the highest scores while English speakers had the lowest scores.

1 Introduction

This paper describes our participation in the interactive track (ChiSwe Group) of the Social Book Search (SBS) lab organized at CLEF 2016¹. This is our first participation in CLEF SBS interactive track. Our group conducted all the experiments in China and the participants' native language in our group is Chinese. It has been of recent interest to explore the search behavior of multi-lingual users [1]. We noticed using the data set from CLEF SBS 2015, Skov and Bogers examined the differences in search behavior between native speakers and non-native speakers of English [2]. But surprisingly, their results showed few significant differences in search behaviors between native and non-native speakers. Other researchers had found that searching in a foreign language requires significantly longer time, more query reformulations, and more websites viewed [3,4]. For example, Hansen and Karlgren concluded that their hypotheses included results for assessment in a foreign, albeit near-native competence, language would be more time-consuming and taxing than those for the first language. Assessing texts in English (27 seconds average assessment time per document) took longer than for Swedish (20 seconds) ($p > 0:95$; Mann Whitney U) [5]. During the experiments in China, our participants have

¹ <http://social-book-search.humanities.uva.nl/#/interactive>

commented that they have difficulties in using the English interface of the search systems and understanding the requirements of search tasks. Since this is the first time that a significant number of Chinese searchers participated in this experiment and English is the language used for the search system, we think it might be interesting to explore what role the native language plays in interactive social book search. Therefore, we divided all the participants in three groups: Chinese native speakers, English native speakers and native speakers of other languages.

Our main research question is: What are the influences of native languages on search behaviors and search experience? Specifically, we will examine four specific RQs:

RQ1. What are the influences of native languages on the task completion time?

RQ2. What are the influences of native languages on the number of search interactions, e.g. book search, browsing and bookbag behavior?

RQ3. What are the influences of native languages on the perceived usefulness of different search tools?

RQ4. What are the influences of native languages on users' engagements during search?

2 Methodology

A total of 112 participants participated in this year's SBS interactive track experiment. The search log data and questionnaire data include participants' demographic information, search activity log, and answers to the questionnaires regarding search experience and engagements. We divided the participants into three groups according to their mother tongues, and the distributions of the number of participants by their native languages are as follows: 29 native Chinese speakers 27 native English speakers, and 56 native speakers of other languages. In order to answer our research questions, we analyzed participants' answers to the questionnaires and the activity log data. In each experiment, the participants were required to perform at least one search task (task 1), and they also had the option to perform a second task (task 2). For task 1, two types of search tasks were designed in SBS interactive track to investigate the impact of task types on the participants' search behaviors in social book search: focused and open tasks. For search behaviors and usefulness judgments on different search tools, we first compare the differences by three groups of searchers for all search tasks; and then in order to compare the differences between two types of tasks, we only selected task 1 for analysis, and compare the differences by three groups of searchers under each type of tasks.

3 Results

3.1 Search time

We first compared the task completion time among three groups of participants. The tests of

normality of task completion time in task 1 and task 2 showed neither of them was normal distributed, therefore, we conducted Kruskal-Wallis Tests on task completion time. Results show that there were significant differences among participants with different native languages ($p < 0.001$ for both task1 and task 2) as shown in Table 1.

Table 1. Comparison of task completion time among three groups of participants (by task)

Tasks	Task completion time Median of each group (in minutes)			Comparison (p value)
	Chinese	English	Other	
task 1	17.79	8.77	8.41	<0.001
task 2	7.23	3.60	3.92	<0.001

Then post-hoc analysis was conducted to compare the differences among them. As shown in Figure 1, for both task 1 and task 2, Chinese speakers spent significantly longer time to complete the tasks than the other two groups of participants. On average, Chinese participants spent 17.79 minutes to complete task 1 on average, whereas English participants spent 8.77 minutes and other language participants spent 8.41 minutes. For task 2, Chinese participants spent 7.23 minutes to complete on average, whereas English participants spent 3.6 minutes on average, and other language participants spent 3.92 minutes on average.

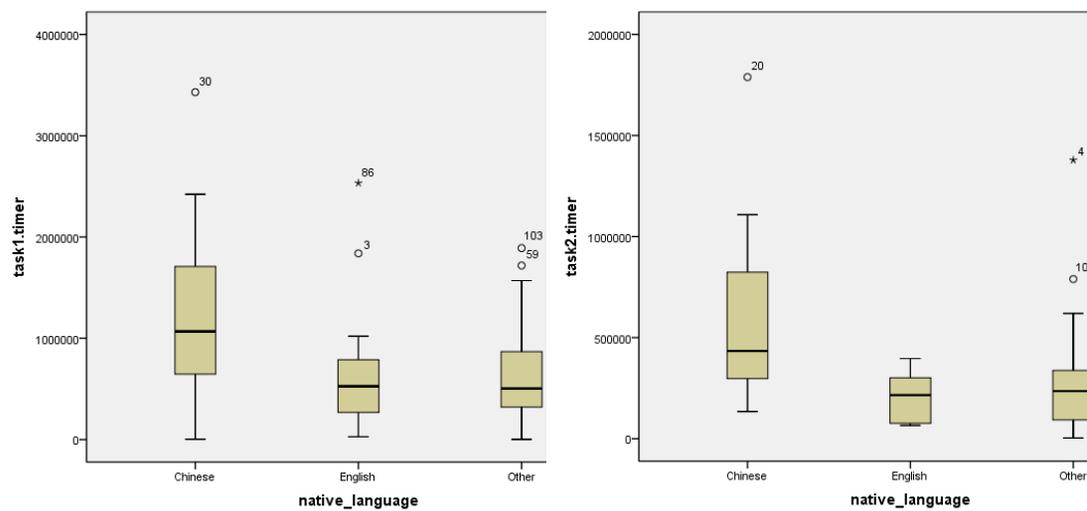


Figure 1. Boxplot for three groups of users in task completion time, for task1 (left) and task2 (right)

Since only task 1 contains task type information, we then focused on task 1 to further examine whether different language participants have any differences in task completion time in each of the two task types.

Table 2. Comparison of task completion time among three groups of participants (by task type)

Task type	Task completion time Median of each group (in minutes)			Comparison (p value)
	Chinese	English	Other	
Focused tasks	24.62	10.56	10.16	0.011
Open tasks	13.42	2.61	6.90	0.03

When only focused tasks were considered, users' completion time also showed significant difference among three groups of participants ($p=0.011$), as shown in Table 2.

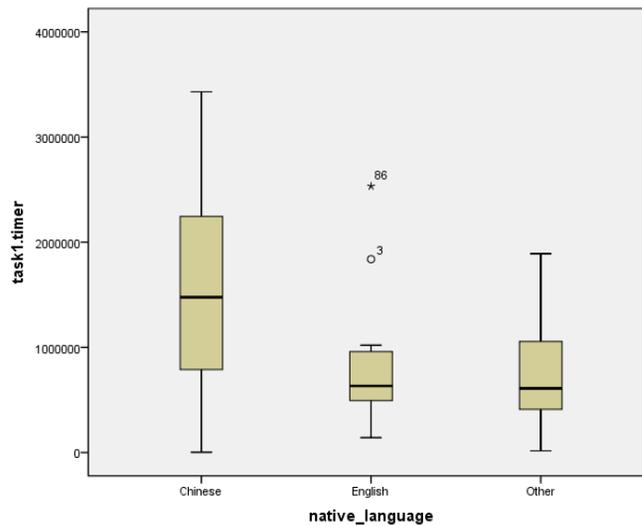


Figure 2. Boxplot for three groups of users in task completion time for “focused tasks”

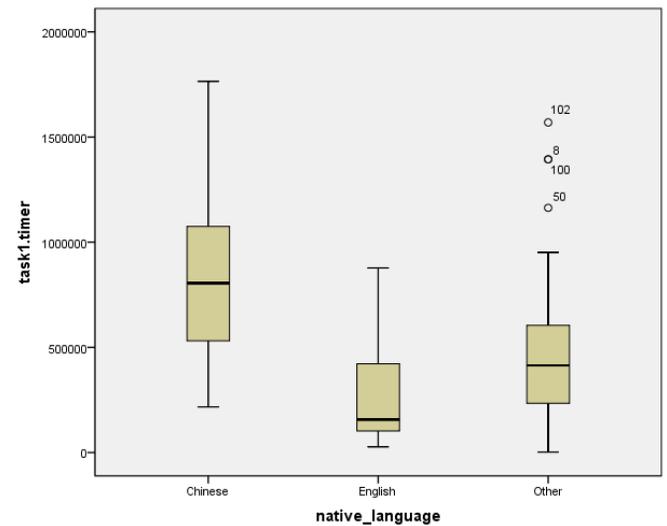


Figure 3. Boxplot for three groups of users in task completion time for “open tasks”

The post hoc analysis (Figure 2) showed that native Chinese speakers (Median=24.62 min) had significantly longer time completion time than native English speaker (Median=10.56

min) and other language speakers (Median=10.16 min). When only open tasks were considered, users' completion time also showed significant difference among three groups of participants ($p=0.03$). The post hoc analysis (Figure 3) showed that native Chinese speakers had significantly longer time completion time (Median=13.42 min) than native English speakers (Median=2.61 min) and other language speakers (Median=6.9 min).

3.2 Number of Interactions

From the activity log data, we extracted the following indicators of users' book search, browsing and annotation behavior:

- Task level: number of books selected, switching between the above modes (showlayout)
- Search: number of queries issued, reset search
- Browsing: browse, add a facet, remove a facet, show item, view metadata, similar books, paginate (next page)
- Book bag: add to bookbag, remove from bookbag, number of books selected, annotate item.

For each task, we extracted the number of times a user performed the above activities. We tested the distribution of the above indicators, and results show that none of them are normally distributed. So we conducted K-Wallis tests to see if any of the above variables shows significant differences across the three language groups (native Chinese speaker, native English speaker, and native speaker of other languages). Table 3 shows a summary of the test results:

Table 3. K-Wallis H Test Results of search behaviors by Language Groups

		Chi-Square	df	Sig.
Task	numbook	14.303	2	0.001**
	showlayout	7.299	2	0.026*
Search	Query	5.1	2	0.078
	resetsearch	1.49	2	0.475
Browsing	Browse	0.405	2	0.817
	addfacet	0.819	2	0.664
	removefacet	4.738	2	0.094
	showitem	8.186	2	0.017*
	metadata	14.745	2	0.001**
	similarbooks	7.187	2	0.028*
	paginate	1.929	2	0.381

Bookbag	addtobookbag	4.834	2	0.089
	removefrombookbag	5.401	2	0.067
	annotateitem	8.781	2	0.012*

3.2.1 Task level comparison

Number of books selected. Results show that there is a significant difference between the three language groups in terms of how many books they selected for each task.

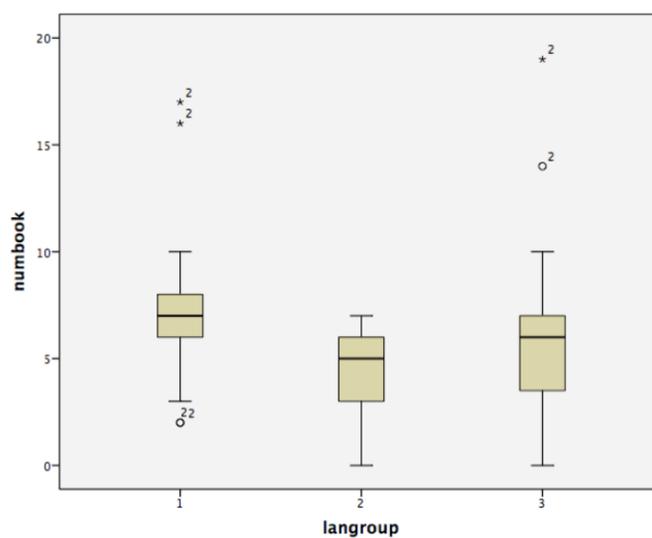


Figure 4: Number of books selected by Language Group (1: Chinese; 2: English; 3: Other)

Figure 4 shows that the Chinese group selected the most number of books, while the English group selected the least number of books for the tasks.

Switching between search, browsing and review. Results show that there is a significant difference between the three language groups in terms of how many times they have switched between search, browsing and review modes.

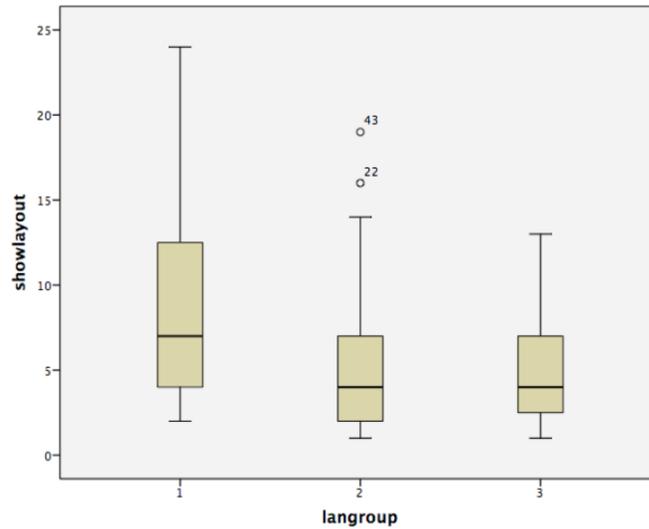


Figure 5: Number of Switching Layout by Language Groups (1: Chinese; 2: English; 3: Other)

Figure 5 shows that the Chinese group seemed to have switched most between search, browse and review mode of the system whereas the other two groups seemed similar.

3.2.2 Search

Results showed that there was no significant difference across three language groups in terms of number of *queries issued* and *number of times users reset search*. Browsing and bookbag activities seemed more different among the three groups.

3.2.3 Browsing

Show item and view metadata. Results show that there is a significant difference between the three language groups in terms of how many times they viewed the metadata of a book.

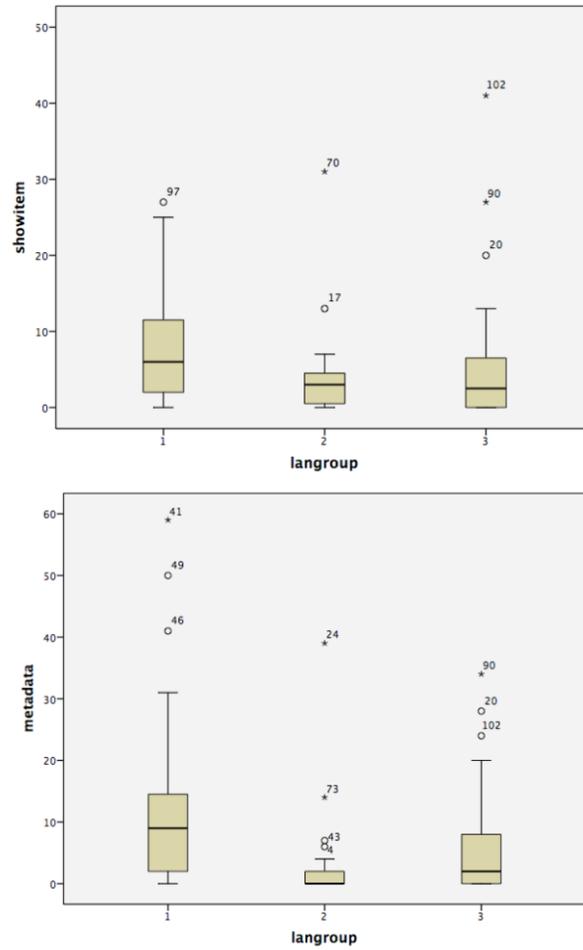


Figure 6: Number of Item and Metadata Viewing by Language Group (1: Chinese; 2: English; 3: Other)

Figure 6 shows that the Chinese group viewed more items and more metadata than the other two groups. There is no significant difference in other browsing activities.

3.2.4 Bookbag Use

Book annotation. Results show that there is a significant difference between the three language groups in terms of how many books they annotated.

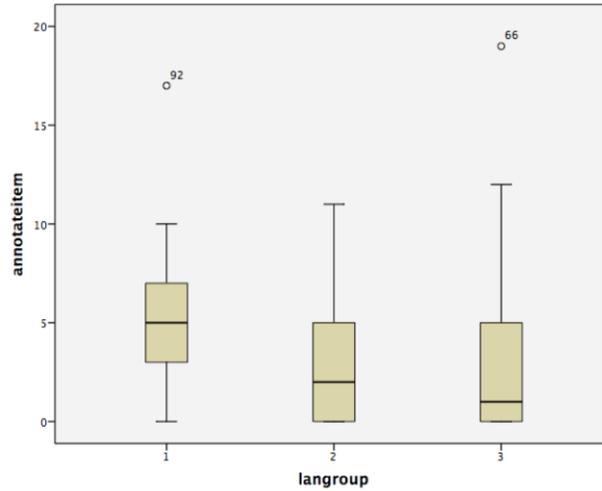


Figure 7: Number of Book Annotation by Language Group (1: Chinese; 2: English; 3: Other)

Figure 7 shows that the Chinese group annotated more than the other two groups, while the other group annotated least number of books (although they selected more books than the native English speakers).

3.2.5 Task Types

We also compared task types (open vs. focused) in addition to native language groups. Table 4 shows the results.

Table 4. K-Wallis H Test Results of search behaviors by Language Groups in two types of tasks

		Focused Task			Open Task		
		Chi-Square	df	Sig.	Chi-Square	df	Sig.
Task	numbook	13.199	2	0.001	12.721	2	0.002
	showlayout	9.538	2	0.008	9.538	2	0.008
Search	query	7.4	2	0.025	9.623	2	0.008
	resetsearch	1.05	2	0.592	2.39	2	0.303
Browsing	browse	2.075	2	0.354	0.747	2	0.688
	addfacet	2.014	2	0.365	1.063	2	0.588
	removefacet	0.02	2	0.99	4.507	2	0.105
	showitem	4.113	2	0.128	4.766	2	0.092
	metadata	7.653	2	0.022	8.454	2	0.015

	similarbooks	3.106	2	0.212	9.717	2	0.008
	paginate	0.786	2	0.675	2.345	2	0.31
Bookbag	addtobookbag	4.92	2	0.085	5.57	2	0.062
	removefrombookbag	1.258	2	0.533	4.34	2	0.114
	annotateitem	1.389	2	0.499	3.727	2	0.155

The results seemed to show that there is a different pattern in terms of number of queries issued by each language group for focused task and open task. Figure 8 shows the results.

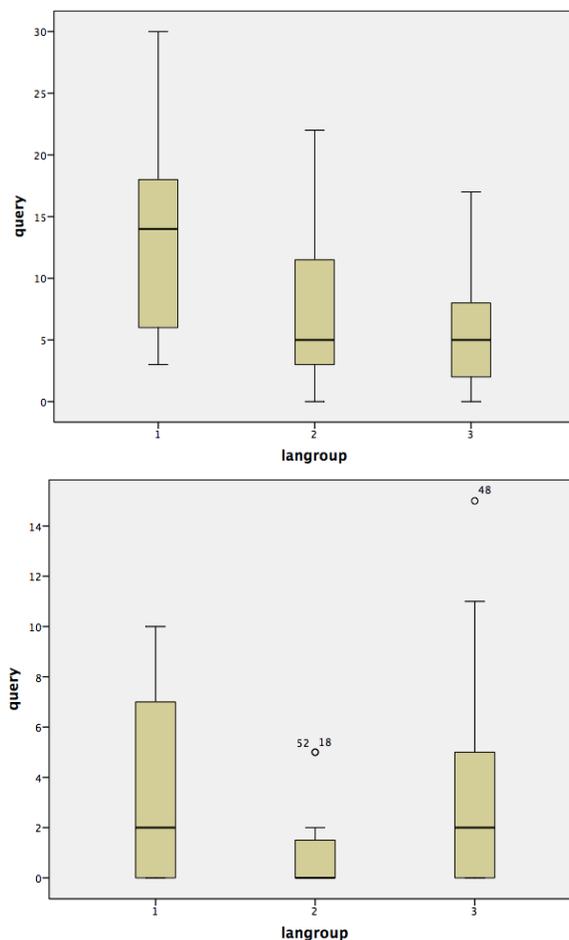


Figure 8: Number of Queries for Focused (left) and Open (right) Tasks by Language Group
(1: Chinese; 2: English; 3: Other)

For the focused task, Chinese users issued a lot more queries than the other group whereas English and other speakers issued similar number of queries. For the open task, English speakers issued least number of queries whereas Chinese and other speakers issued similar number of queries.

For the open task, Chinese users seemed to use more “similar books” feature, relying on system recommendation whereas the other two groups did not use this feature as much.

3.3 Users’ Perceptions of the Usefulness of Search Tools

After participants completed each of search tasks, there was a post-task questionnaire to ask about the usefulness of each search tools used during searching. In this part, we compared participants’ judgments of the usefulness among three native language groups. Since we are interested to see if there is any difference between focused and open tasks, we only focused in users’ evaluations in task 1 in this part. Similarly, Kruskal-Wallis Tests were conducted for task 1, and then specifically for focused and open tasks.

As shown in Table 5, when searching for task 1, users have significant differences in two tools among three groups: browse individual books and search results. The post-hoc demonstrated that English searchers rated browse individual books significantly lower than other language searchers, and Chinese searchers were not significantly different with the other two groups of searchers, as shown in Figure 9. For search results page, the post-hoc showed that English searchers rated significantly lower than the other two groups of searchers, and there is no significant difference between Chinese and other language searchers, as shown in Figure 10.

Table 5. K-Wallis H Test Results of the usefulness of search tools by Language Groups

Search tools	Kruskal-Wallis Tests (p value)		
	Task 1	task=focused	task=open
bookbag.notes	0.677	0.963	0.543
bookbag.similar_books	0.771	0.212	0.597
browse.individual_books	0.049	0.161	0.073
browse.topic_explorer	0.398	0.918	0.371
meta_data.description	0.204	0.040	0.113
meta_data.publication	0.479	0.152	0.892
meta_data.reviews	0.054	0.257	0.350
meta_data.tags	0.665	0.323	0.206
search.search_box	0.084	0.039	0.777
search.search_facets	0.633	0.246	0.719
search.search_history	0.592	0.607	0.904

search.search_results	0.007	0.063	0.120
search.search_topic	0.725	0.373	0.440

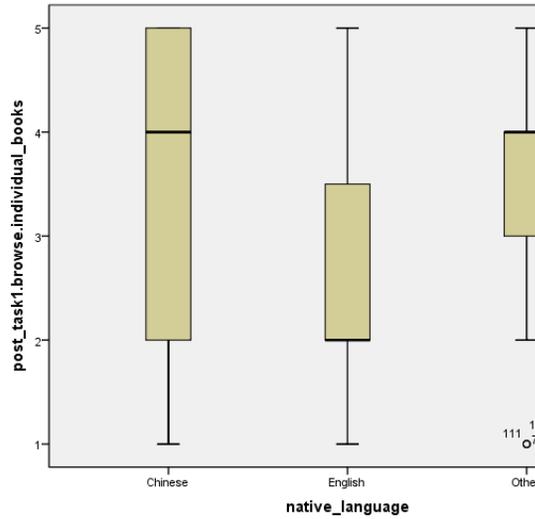


Figure 9. Boxplot for three groups of users in the usefulness of browsing individual books

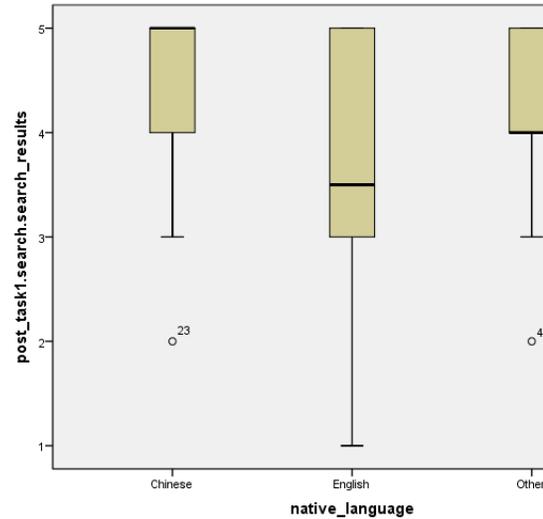


Figure 10. Boxplot for three groups of users in the usefulness of search results

When only “focused” tasks were considered, two tools showed significant differences: meta.data.description, and search.box. For meta.data.description, the English searchers rated significantly lower than the other two groups, as shown in Figure 11. For search box, Chinese searcher rated significantly more useful than the other two groups of searchers, as shown in Figure 12.

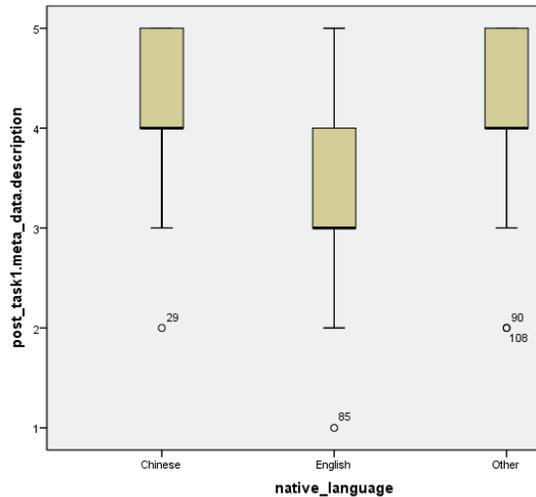


Figure 11. Boxplot for three groups of users in the usefulness of meta.data.description

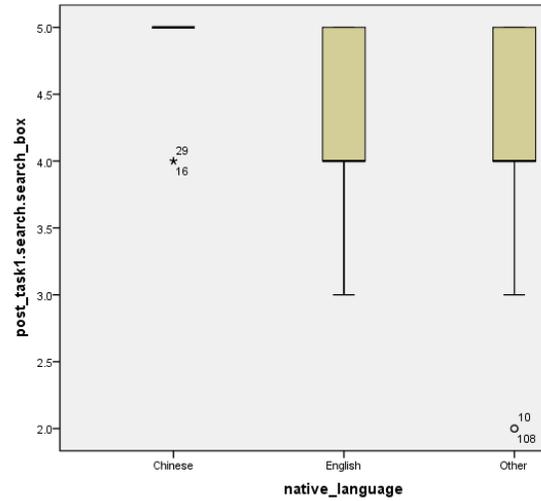


Figure 12. Boxplot for three groups of users in the usefulness of search_box

3.4 Search Engagements

After participants had completed both search tasks, they were asked to fill out a questionnaire about their engagements for the search system. The engagement questionnaire consisted of 31 questions representing six groups of engagement factors: aesthetics, durability, focused attention, felt involvement, novelty, and perceived usefulness. Since the website was designed in English language only, we could hypothesize that participants with different native languages, especially whether native English speakers, native Chinese speakers and other speaker had engaged in searching using this system at different levels, due to different language (English) proficiency levels.

First of all, we tested the normality of the engagement variables, and found none of them were normal distributed. Therefore, for the comparison among three groups of participants, we used Kruskal-Wallis Tests. The results are shows in Table 6. Among all 31 engagement items, 7 of them were found to be significantly different among the three groups of participants: **en1** (*Exploring this website was worthwhile*), **en4** (*My exploration experience was rewarding*), **fa4** (*When exploring, I lost track of the world around me*), **fa5** (*The time I spent exploring just slipped away*), **fa6** (*I was absorbed in exploring*), **fi1** (*I was really drawn into my exploration task*), **pu1** (*I felt frustrated while exploring this website*).

Table 6 K-Wallis H Test Results of engagements by Language Groups

Factors	Variables	Items	Chi-Square	Sig.
Aesthetic	ae1	This website is attractive	3.166	0.205
	ae2	This website was aesthetically appealing	2.028	0.363
	ae3	I liked the graphics and images used on this websites	0.257	0.880
	ae4	This website appealed to my visual senses	0.523	0.770
	ae5	The screen layout of this website was visually pleasing	2.969	0.227
endurability	en1	Exploring this website was worthwhile	9.119	0.010
	en2	I consider my experience a success	2.422	0.298
	en3	This experience did not work out as I had planned	1.347	0.510
	en4	My exploration experience was rewarding	10.401	0.006
	en5	I would recommend exploring this website to my friends and family	4.453	0.108
Focused Attention	fa1	I lost myself in this experience	2.185	0.335
	fa2	I was so involved in this experience I lost track of time	3.939	0.140
	fa3	I blocked out things around me when I was exploring this website	4.921	0.085
	fa4	When exploring, I lost track of the world around me	6.436	0.040
	fa5	The time I spent exploring just slipped away	7.247	0.027
	fa6	I was absorbed in exploring	9.796	0.007

	fa7	During this experience I let myself go	1.942	0.379
felt involvement	fi1	I was really drawn into my exploration task	14.280	0.001
	fi2	I felt involved in this exploration task	3.732	0.155
	fi3	This exploration experience was fun	5.284	0.071
Novelty	no1	I continued to explore this website out of curiosity	2.848	0.241
	no2	The content of the website incited my curiosity	1.296	0.523
	no3	I felt interested in my exploration task	2.807	0.246
perceived usability	pu1	I felt frustrated while exploring this website	8.805	0.012
	pu2	I found this website confusing to use	4.826	0.090
	pu3	I felt annoyed while visiting this website	2.412	0.299
	pu4	I found this website confusing to use	0.386	0.825
	pu5	Using this website was mentally taxing	2.267	0.322
	pu6	this experience was demanding	0.647	0.724
	pu7	I felt in control of my exploration experience	2.508	0.285
	pu8	I could not do some of the things I needed to do	1.691	0.429

We then conducted Bonferroni tests for post-hoc analysis for pairwise comparisons.

For en1 (*Exploring this website was worthwhile*), the post-hoc analysis showed native English speakers were significantly different from both Chinese and other language speakers. Particularly, English speakers rated significantly lower (M=1.67) than Chinese speakers (M=2.61) and other language speakers (M=2.34).

For en4 (*My exploration experience was rewarding*), the post-hoc analysis showed native English speakers (M=1.63) rated significantly lower than Chinese speakers (M=2.68), and other language speakers were not significantly different from English or Chinese speakers

(M=2.18).

For fa4 (*When exploring, I lost track of the world around me*), the post-hoc analysis showed native English speakers (M=1.07) rated significantly lower than Chinese speakers (M=1.86), and other language speakers were not significantly different from English or Chinese speakers (M=1.32).

For fa5 (*The time I spent exploring just slipped away*), the post-hoc analysis showed native English speakers (M=1.41) rated significantly lower than Chinese speakers (M=2.25), and other language speakers were not significantly different from English or Chinese speakers (M=1.59).

For fa6 (*I was absorbed in exploring*), the post-hoc analysis showed other language speakers (M=1.86) rated significantly lower than Chinese speakers (M=2.79), and English speakers were not significantly different from Chinese or other language speakers (M=2.11).

For fi1 (*I was really drawn into my exploration task*), the post-hoc analysis showed Chinese speakers were significantly different from both English and other language speakers. Particularly, Chinese speakers rated significantly higher (M=2.64) than English speakers (M=1.48) and other language speakers (M=1.89).

For pu1 (*I felt frustrated while exploring this website*), the post-hoc analysis showed Chinese speakers (M=1.21) rated significantly lower than English speakers (M=2.22), and other language speakers were not significantly different from English or Chinese speakers (M=1.57).

4 Discussion and Conclusion

This notepaper presents the preliminary results on the influences of native languages in search behaviors and search experiences in the context of interactive social book search. Earlier studies in the proceeding of CLEF SBS 2015 did not find many differences between native and non-native speakers of English. This year, we joined in CLEF SBS interactive track, and we think it is reasonable to examine the difference among native Chinese speakers with native English speakers and other European languages. In general the results show a series of differences among the three groups of participants. The results show that Chinese searchers devoted more search efforts in searching, i.e. spent longest time to complete search tasks, selected the most number of books, have switched most between search, browse and review model of the search system, viewed more items and more metadata, and annotated more than English and other language speakers. This is consistent with the results in [3]. Comparatively, English searchers had spent least search efforts among the three groups of searchers. Besides language effects, another possible reason for this is that all the Chinese speakers conducted searching in the lab mode in this experiment, and all other participants conducted searching remotely. Since we do not have any Chinese speakers who participated remotely, there is no way for us to filter out the effect of participation modes in this analysis. We should consider having more Chinese participants remotely in the future to further validate this result. With respect to the usefulness of search tools, few significant differences were found. For the two search tools that showed significant differences, i.e.

browse_individual_books and search_results, English speakers had the lowest score of usefulness among three groups of searchers. For the engagement comparison, seven measures were found to be significantly different among three groups of searchers. In general, English searchers had the lowest score in the engagement with the search system, and Chinese searchers had the highest score in the engagement. We should further explore the data to explain such phenomena. One possible explanation is that Chinese searchers have devoted most efforts in searching, so they tended not to rate the system with the lowest measure.

References

1. Steichen, B., Ghorab, M. R., O'Connor, A., Lawless, S., & Wade, V. (2014). Towards Personalized Multilingual Information Access-Exploring the Browsing and Search Behavior of Multilingual Users. In *User Modeling, Adaptation, and Personalization* (pp. 435-446). Springer International Publishing.
2. Bogers, T., Gäde, M., Hall, M. M., & Skov, M. (2016). Analyzing the influence of Language Proficiency on Interactive Book Search Behavior. *Proceedings of Iconference 2016*.
3. Chu, P., Jozsa, E., Komlodi, A., & Hercegfı, K. (2012, August). An exploratory study on search behavior in different languages. In *Proceedings of the 4th Information Interaction in Context Symposium* (pp. 318-321). ACM.
4. Rózsa, G., Komlodi, A., & Chu, P. (2015, May). Online Searching in English as a Foreign Language. In *Proceedings of the 24th International Conference on World Wide Web Companion* (pp. 875-880). International World Wide Web Conferences Steering Committee.
5. Hansen, P. and Karlgren, J. (2005). Effects of foreign language and task scenario on relevance assessment. *Journal of Documentation*, Vol. 61 (5), 2005, pp. 623-638.