

THE IMPACT OF LANGUAGE ON WEBSITE USE AND USER SATISFACTION: PROJECT DESCRIPTION

Kralisch, Anett, Humboldt-Universität zu Berlin, Institut für Wirtschaftsinformatik,
Spandauer Str. 1, 10178 Berlin, Germany, kralisch@wiwi.hu-berlin.de

Köppen, Veit, Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft, Garystr. 21,
14195 Berlin, Germany, koeppen@wiwiss.fu-berlin.de

Abstract

This paper describes ongoing research investigating the ways in which language affects both the way users seek and find information on a website, and their eventual level of satisfaction with the site's offering. Our aim is to understand how language may form a barrier to accessing information, and to identify ways of overcoming these boundaries. We build on the Theory of Information Foraging, which predicts web-page access based on the benefit compared to the costs of accessing additional information.

We identify aspects of language that determine the costs and values of a search, thus determining a website's language-related information value. Behavioural and attitudinal outcomes of perceived information value - website use and satisfaction - are discussed. We formulate hypotheses and a model that are currently empirically evaluated. Advantages of this model are a theoretically grounded derivation of core determinants of perceived information value, and a demonstration that data describing these determinants can be inferred from data that are easily accessible over the Internet.

Expected results will provide support for investment decisions concerning website translations and adaptations. Such design decisions can help improve access and participation among citizens from different linguistic backgrounds, thus reducing digital divides arising from language-related factors.

Keywords: Multilingual websites, Website Translation, Information Foraging Theory, Information value.

1 INTRODUCTION

Given the number of languages spoken in the world, a comprehensive native-language website is both technologically and economically infeasible in all markets. As a consequence, Internet users often surf on websites that present information in languages other than their native tongues. This gives rise to a number of questions: Do users' proficiency levels have a direct effect on website use and user satisfaction? Are there other reasons for using (or not using) a non-native-language website and feeling comfortable with it? Moreover, since a comprehensive translation of a website is not feasible, decision support should be provided for choosing the languages to be considered for translation.

Answers to these questions are helpful for two reasons. The first is commercial: since website translations and adaptations often represent important investment decisions, data and knowledge about users' language-related decision processes are valuable for appropriate linguistic adaptations. The second reason is ethical: the goal of increasing participation in Internet communication, of reducing the "digital divide". This perspective is gaining importance with the increasing desire to enable wide-ranging citizen participation on the Internet.

The theoretical contribution consists of investigating the ways in which language affects both the way users seek and find information on a site, and their eventual satisfaction with the site's offered result. The research described aims to detect the impact of language as a boundary in accessing information and to identify ways of overcoming these boundaries. We build on the Theory of Information Foraging. We propose a model that makes predictions from two types of data: (a) Internet-wide data that are open and easily accessible, and (b) site-specific data that are obtained from our analyses.

2 RELATED RESEARCH

Pirolli and Card (1999) proposed the Theory of Information Foraging (IFT) as an explanation of user strategies for information gathering on the Internet. According to the model, users continue to follow links as long as the information gained from following the link is not exceeded by the cost of accessing it. The model permits the prediction of web users' flow and webpage access, and to infer users' information needs through the application of the reverse form of the model (Chi et al. 1999). Predictive models that are based on the IFT are intrinsically complex. The number of available links, the number of previously accessed pages, the search goal, and numerous other determinants were determined as having an impact on the perceived value of information gained (Bernard 2000).

So far, language-related aspects have not been considered in studies of Information Foraging. However, it is well known that language proficiency strongly affects reading and information access (e.g. Chebat et al. 2003). Therefore, we aim to investigate the impact of language on perceived costs and values of information gathered on the Internet. In addition to language *knowledge*, the *perception* of information offered in foreign languages will be taken into account as a second linguistic variable. We thereby extend other analyses regarding the impact of language. In the second step of our analysis, we investigate the impact of the language-related cost and value of information gathering on two important indicators of website success: use and satisfaction.

The paper is structured as follows: First, we define a website's *language-related net value*. We identify aspects of language that can be considered as costs and benefits of an information search, and hence as antecedents of a website's language-related information value. Behavioural and attitudinal outcomes of perceived information value are discussed in Section 3.3. We specify hypotheses that are currently evaluated empirically. Section 4 describes the research method. We conclude by providing preliminary results, discussing the limitations of our study and giving an outlook on future research and on applications of the expected results.

3 CONCEPTUAL FRAMEWORK AND HYPOTHESES

3.1 The perceived language-related net value of a website

The term “net value” denotes the value of gathered information after deducting the cost of collecting it. Previous research has mostly described the consequences of perceived net value (such as link following) rather than the value itself. We introduce net value as a perceptual construct because it depends primarily on the user’s *subjective evaluation* of language costs and values. It includes perceptions of advantages and disadvantages of native and non-native language offerings, as well as specific attitudes towards a non-native language offering.

3.2 Antecedents of a website’s language-related net value: language proficiency and Internet language offer

The IFT regards link-following behaviour as a resource allocation problem. For the purpose of our investigation we focus on the allocation of two language-dependent resources: cognitive effort and time. Cognitive resources are required for language processing while gathering information. The Revised-Hierarchy-Model by Dufour and Kroll (1995) explains why language processing in a non-native language (L2) requires more effort than language processing in a native language (L1). For non-native-language users, the degree of effort is determined by their L2 proficiency. From this we infer:

- H1: A user’s perception of a website’s language-related net value increases with his/her language proficiency.

Along with users’ linguistic knowledges, the overall Internet presence of a language is also expected to have an impact on perceived net value. Economic theory predicts that a product’s value decreases with an increase in the supply within a market (for details see e.g., Grin 1996). Transferring this assertion into language issues with the Internet market leads us to expect that the marginal language-related value of a web page decreases as more pages are offered in that language on the Internet¹. As a further consequence, users with a large amount of web offerings in their native tongue are more likely to perceive a lower value of a non-native language website. This leads to the following hypothesis:

- H2: A user’s perception of a website’s language-related net value decreases with an increase in the perceived amount of information offered on the Internet in the site’s language.

Data on both language proficiency and Internet languages offered are easily accessible on the Internet or in the literature (see 4.). To increase data reliability, we will compare (a) the self-report data users provide in a questionnaire with the official language(s) of the country from which they access the site, and (b) users’ self-reported perceptions of Internet language distribution with global Internet statistics..

3.3 Outcomes of the language-related perceived net value of websites: use of and satisfaction with the site

Website usage and website satisfaction are two variables that are considered to be important indicators of success. It is therefore important to understand how a site’s language affects both usage and satisfaction. The relationship between expected information value and website access has been demonstrated in previous IFT studies. In contrast to those studies, we concentrate on navigation behaviour *within* one website rather than *across* various websites. Since both kinds of analyses focus on link-following behaviour, this difference can be disregarded.

¹ This approach represents a simplified view on a language’s value on the Internet. Other aspects such as link distribution/network externalities are likely to also have an impact. However, we consider these factors to have a smaller impact on users’ perceptions than the size of the offer in a given language.

We propose regarding information evaluation as a two-step process. In the first step, based on proximal cues and inferred information value, the user decides whether or not to follow a link. If the access costs do not exceed the expected value, the link is followed. Perceiving language as a cue, a website's offered languages should consequently have an impact on which web pages are accessed by which users. This suggests that a site's language(s) also affect access patterns within the site, such as *how many* and *which* pages are accessed (see also Chi et al. 1999). We expect:

- H3: The number of pages accessed by a user on a site to increase with the perceived language-related net value of that site.

Extending previous research in Information Foraging, we propose to take into account a second evaluation step of a web page's value. This step occurs *after* the page is accessed. Here the user decides to what extent the content of the web page is worth being processed (for a two-step decision process see also Häubl & Trifts 2000). We propose the following:

- H4: The depth of processing a user performs during a visit to a site increases with the perceived language-related net value of that site.

In addition to the investigations of behavioural outcomes, the relationship between perceived net value and perceived satisfaction is also examined. We expect that:

- H5: User satisfaction increases with perceived net value.

The informational value of a website can be thought of as its perceived "usefulness", and the cost of information access can be thought of as the opposite of the site's "ease of use". Usefulness and ease of use have been shown by Davis (1993) to be significant determinants of the acceptance of technology, in terms of attitudes toward the system and actual system use.

Based on the hypotheses stated above, we developed a model of the language-related value of a website (see fig. 1). Due to the constructs employed and the relationships postulated, it can be regarded as an instance – or a facet – of Davis' general model of technology acceptance.

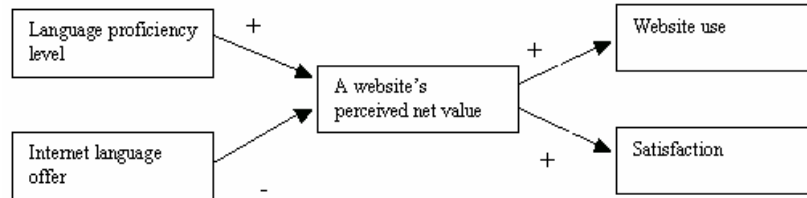


Fig. 1: The language-related website value model. Hypothesized relations are indicated according to the study's research propositions. "+" indicates a positive relationship, "-" negative.

3.4 Control variables

Besides demographic data, we consider medical knowledge, internet experience, reason for visiting the website, and the required language processing effort of a web-page as important background variables to be tested (for reasons for eventual relationships, see Carrell & Wise 1998, Steffenson et al. 1979, Hölscher & Strube 2000, Bernard 2000).

4 DATA AND METHOD

4.1 Data

Data is obtained through a combined analysis of the logfile from a largely frequented multilingual e-Health website and a multilingual questionnaire posted there. Previous investigations have shown that an extraordinarily large linguistic and cultural variability among the users can be expected (Kralisch &

Berendt 2004 a, b). Additionally, data from Internet statistics (Cyber Census Report, Mikami & Suzuki 2004; and the DMOZ.Org World Statistics) and countries' official language statuses are collected and integrated into a language database. We augment this information using software developed in our group (a) for measuring web page complexity by a self-developed index based on the amount of text, pictures, and outgoing hyperlinks in that page (see also Waddil & McDaniel 1992), and (b) for inferring geographical information from IP addresses and referrers.

Session-IDs allow us to link questionnaire data to logfile records of site usage. The logfile includes information about the user's IP address, the requested page, the time of access, the language of the requested page, and the page that was requested before (referrer page). It will be pre-processed according to the standards of web usage mining (e.g., Cooley et al. 1999). Since the logfile is linked to the questionnaire, problems with sessionizing or robots will not occur.

4.2 Measures

Information about the users' language proficiencies and website language use are self-reported by participants in the questionnaire. The questionnaire is complemented by our language database. Users' answer on their language choice on the website can be verified through data from the logfile. The users also rate, on a 7-point-Likert scale, how common they consider a website offer in their native tongue. The questionnaire furthermore asks the users about their beliefs regarding how the language offered affect the users' perceived net values in terms of possible time-savings, beliefs about their extent of website use², comprehensibility and accessibility, etc. The users are also asked to evaluate, through multi-item-scales, various aspects of satisfaction received from website use. The scales are either Likert scales or semantic differential scales with a range of seven points each.

The objective measurement of the extent of website use is mainly assessed by means of logfile analysis. We will measure the number of requested pages and the time spent on the pages. Since an exhaustive psycholinguistic investigation into the depth of processing a page would go far beyond the purpose of our study, we examine the degree of content processing simply in terms of time spent on a page. Due to higher cognitive effort and longer content processing times, we expect users of lower language proficiency levels to choose only a small number of pages for extensive content processing. Those pages that are requested but not selected for content processing are expected to be scanned very briefly. As a consequence, navigation behaviour should be characterized by high standard deviations of time spent on web-pages.

In order to test our hypotheses and identify the strength of the impact of language we will conduct confirmatory factor analyses and test the correctness of our structural model using LISREL. By the means of these analyses we will compute statistical aspects such fit indices, factor loadings, intercorrelations among the study's model and control variables, correlations among the model's variables and the strength of path coefficients (see also Bollen 1989, Chin 1998).

4.3 Preliminary results

We were able to obtain preliminary results for a first testing of our model from the data of a pretest-questionnaire. Due to the small sample size (N=135), we refrained from employing software-based structuring equation modeling. Instead, only path coefficients between variables were calculated. Among the survey participants 83.7% were native speakers of one of the website's languages and 16.3% were non-native speakers. This ratio confirms results obtained in a previous study of the same website (Kralisch & Berendt 2004a). Examinations of the control variables revealed a highly

² We assess the belief regarding the extent of website use (part of perceived net value) as well as the actual usage patterns (part of website use) in order to detect discrepancies between beliefs and actual behaviour. Inconsistencies between these two variables have been described for example by Nisbett & Wilson (1977).

significant correlation between medical knowledge and native tongue (L1 vs. L2) ($r=.332$). Avoiding a more complex model, we split the sample into two groups. Since the physician sample did not provide sufficient empirical data, all the following results refer to the patients group only.

Fig. 2 summarizes the discovered path coefficients. Path coefficients here are standardised partial regression coefficients (b-weights) (Bortz 2005). We used backward regression analyses with standardised variables to test whether more variables than those predicted in the model considerably affect the dependent variable. Calculations are weighted by the number of L1 and L2 speakers.

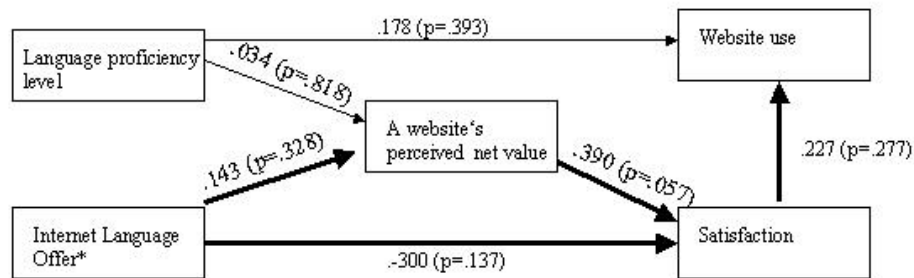


Fig. 2: Results language-related website value model. A \longrightarrow indicates the variable with the strongest impact, a \longrightarrow indicates the variable with the second strongest impact.

* the "Internet language offer" is coded in the following way: low number = large offer; high number = small offer.

None of our hypotheses can be confirmed since none of the results are significant. However, non-significance might be to a certain extent due to the small sample size. Nonetheless, it should be noted that the direction of the correlations (negative vs. positive) confirms in all cases our predictions.

Results indicate that the language proficiency level and the perceived Internet offering affect website use and satisfaction. Yet the impact does not always occur in the expected path. In particular, the intermediate concept of the net value needs to be partly revised: first, the coefficient of determination ($R^2=0,021$) is a sign of a minor explanatory value of language proficiency and internet language offer, second, the impact of the language proficiency level seems to be almost negligible, and third, the net value itself seems not to affect the number of requested pages (Beta=0.006, $p=.980$).

Moreover, results suggest the direct impact of language proficiency and Internet offer on website use and satisfaction. Language proficiency appears to be positively correlated to the number of page requests and users who perceive a larger Internet offer are more satisfied with the website, and vice versa. While the first correlation is intuitive³, the second is more difficult to interpret. The direct impact might indicate that the website's content and design are more adapted to the preferences of L1 speakers than to those of L2 speakers. Such an interpretation would still be in line with the indirect impact of the Internet language offered on satisfaction that was also found. The indirect impact corresponds to H2 and H5 and supports the intermediate concept of perceived net-value. In order to get better insight, we suggest a separate analysis of L1 and L2 speakers with a larger data sample.

It should be noted again that these results are obtained only from a small sample of a pre-test questionnaire, and refer to the patients group only. Data and results are therefore not representative, nor was extensive structural equation modeling employed for analyses. This will be part of analyses with a larger data set. In particular, latent variables with several indicator variables and error terms will be incorporated. Finally, variables and investigations that have not yet been part of these preliminary analyses will be added.

³ A lower language proficiency level increases the cognitive effort and decreases the ease of use. According to the TAM model, ease of use is positively correlated to the system use. Our correlation is therefore in line with the TAM model.

5 LIMITATIONS AND FUTURE RESEARCH

The model presented in this paper only includes some aspects of the impact of language. It only represents a first step towards a complex understanding of the languages' impacts on the perception of information scents, neglecting other aspects of language.

First, we have included neither "market-independent" aspects of language attitude and language motivation (e.g., language fear/self-confidence) nor language-associated values. Language attitude and language motivation have both been shown to affect users' willingness to read and understand foreign languages (Chebat et al. 2003, Clément et al. 1977). Language-associated values have been investigated in marketing and consumer research (e.g., Dmoch 1997). However, the role of language-associated values has not been analysed yet with regard to the medium of the Internet, nor has its impact on information scent been investigated. Moreover, the mere number of languages offered on a website might also be perceived by the consumer as a cue for the website's quality. Further analyses are clearly needed here.

Second, statistical analyses are conducted for this study on an individual level. Nonetheless, we suggest aggregating data in a second step to the national level. The reason for this is that countries usually represent the investment unit on which business decisions are based. Yet analysis on the national level might reveal differences that are due to cultural habits. Thus, a data aggregation to the national level would first require a detailed analysis of the possible impact of culture. In order to avoid aggravating complexity in this first study, we decided to make these issues part of follow-up investigations.

Finally, the study shares the shortcomings of other field studies such as the possible website specificity and the limited control of personal and situational variables. Future research should validate the results through laboratory experiments and hereby provide further insight.

6 IMPLICATIONS OF EXPECTED RESULTS

Linguistic adaptations of websites are often related to important investment decisions. Therefore, websites are frequently translated into a limited number of languages. Companies that want to reach a worldwide audience without any specific focus generally base their decisions about linguistic adaptations on e-commerce volume and growth. However, especially when the company cooperates with partners of different, yet specific linguistic backgrounds, information about national e-commerce volumes is hardly useful for deciding about which language versions should have the highest priority.

In such situations, information about the likely impact of a linguistically adapted website is required. We expect the results of our study to provide a first useful decision support. First, we expect to gain insight into the respective impact of both the user's language knowledge and his or her perception of a foreign language offering on the perceived net value of a site. Second, we expect results concerning the effect of this "linguistic satisfaction" on the overall satisfaction with the site, on its usage, and thus ultimately its success. The use of language on a website can therefore be seen as a marketing instrument that affects consumer behaviour. Lastly, one of the advantages of our model is that the data used for measuring the determinants are generally easy to obtain. Information about official languages, vernacular languages in use, and the distribution of languages on the Internet are accessible online and from other sources.

References

- Bernard, Michael Lewis (2000). Examining a Metric for Predicting the Accessibility of Information within Hypertext Structures. Unpublished Ph.D. Thesis at Wichita State University, Human Factors Psychology.
- Bollen, K.A. (1989). Structural Equations with latent variables, Wiley, New York.

- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*, Springer, Heidelberg.
- Carrell, Patricia L. & Teresa E. Wise (1998). The Relationship between prior knowledge and topic interest in second language reading. *Studies in Second Language Acquisition*, 20, 285-309.
- Chebat, Jean-Charles, Claire Gelinas-Chebat, Sabrina Hombourger, Arch G. Woodside (2003). Testing Consumers' Motivation and Linguistic Ability as Moderators of Advertising Readability, *Psychology and Marketing*, 20 (7), 599-624, July 2003.
- Chi, E.H., Pirolli, P., Pitkow, J. (2000). The scent of a site: A system for analysing and predicting information scent, usage, and usability of a web site. In *Proceeding of the Association for Computing Machinery's CHI'00*, 161-168.
- Chin, W.W. (1998) Issues and opinions on structural equation modeling, *MIS Quarterly* 22(1),7-16.
- Clément, Richard, Robert C. Gardner & Patric C. Smythe (1977). Motivational variables in second language acquisition: A study of francophones learning English, *Canadian Journal of Behavioural Science*, 9(2).
- Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Journal of Knowledge and Information Systems*, 1(1), 5-32.
- Davis, Fred D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioural impacts. *International Journal of Man-Machine Studies*, 38, 475-487.
- Dmoch, Thomas (1997). *Interkulturelle Werbung. Verhaltenswissenschaftliche Grundlagen für die Standardisierung erlebnisbetonter Werbung*, Aachen.
- DMOZ.Org World Statistics. <http://et.com.mx/mgb/dmoz/stat/index.en.html> last visited: 11/12/2004.
- Dufour, R. & Kroll, J. (1995). Matching Words to Concepts in Two Languages: A Test of The Concept Mediation Model of Bilingual Representation. *Memory & Cognition*, 23, 166-180.
- Grin, François (1996). The economics of language: survey, assessment, and prospects. *International Journal of Society and Language*, 121, 17-44.
- Häubl, Gerald & Valerie Trifts (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, 19 (1), 4-21.
- Kralisch, Anett & Bettina Berendt (2004a). Linguistic Determinants of Search Behaviour on Websites. In: *Proceedings of the Fourth International Conference on Cultural Attitudes towards Technology and Communication*, Karlstad, Sweden, 27 June-1 July 2004.
- Kralisch, Anett & Bettina Berendt (2004b). Cultural Determinants of Search Behaviour on Websites, *Proceedings of the Sixth International Workshop on Internationalisation of Products and Systems*, Vancouver, Canada, 8-10 July 2004.
- Mikami, Yoshiki & Izumi Suzuki (2004). The Language Observatory Project and its Experiment: Cyber Census Survey. In *Proceedings of SCALLA 2004, Crossing the Digital Divide*, Kathmandu, Nepal, 5-7 January 2004.
- Nisbett, Richard E. & Timothy DeCamp Wilson (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231-259.
- Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review*, 106 (4), 643-675.
- Steffenson, M.S., Joag-Dev, C. & Anderson, R.C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 15 (1), 10-29.
- Waddill, Paula J. and Mark A. McDaniel (1992). Pictorial Enhancement of Text Memory: Limitations Imposed by Picture Type and Comprehension Skill. *Memory & Cognition* 20(5), 472-482.