Information Retrieval through Communicative Formulae Within the Internet Interaction

Natalya Zavyalova

Ural Federal University (UrFU), 620002, 19 Mira street, Ekaterinburg, Russian Federation Position: Associate Professor

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 8 July 2014
Accepted 10 August May 2014
Available online 30 August 2014

Keywords: corpus, socio-cultural interaction, communication, code, stereotype, information retrieval.

ABSTRACT

The Internet communication is of primary importance in modern world. How does it work, what are the constituent blocks and relevant communicative formulae, are fundamental questions for successful socio-cultural interaction. With the help of the online electronic linguistic corpuses BNC (British National Corpus) and COCA (Corpus of Contemporary American English) I analyze frequency indexes of communicative formulae to retrieve the information about English native speakers’ communicative preferences. The results of my omnibus survey testify to the fact that communicative formulae were accepted as an effective information channel by a large proportion of respondents. The research results have been processed into a database, marked with the Rospatent Certificate № 2013620397, dated March, 3, 2013.

© 2014 AENSI Publisher All rights reserved.

To Cite This Article: Natalya Zavyalova., Information Retrieval through Communicative Formulae Within the Internet Interaction. Adv. Environ. Biol., 8(13), 315-320, 2014

INTRODUCTION

The idea of the Internet as a major channel of communication is axiomatic nowadays. More and more households throughout the world are getting access to the Internet. According to Eurostat (see Fig.1), the access to the Internet is generally on the rise across modern Europe. National differences in levels of access persist, but are decreasing over time.

Fig. 1: Percentage Household Internet Penetration in Europe in 2007 and 2012 [8].

However, according to the statistics of NTIA (National Telecommunication and Information Administration, the USA), there’s still a tangible digital divide across the country: “the results indicate that households with lower incomes and less education, as well as Blacks, Hispanics, people with disabilities, and rural residents were less likely to have home Internet access service” [9].

With the increased role of the Internet access as an indicator of wealth and high social status it is not surprising that electronic media are used more and more to influence public opinions. “Media and communications have always been employed by dominant actors and played a crucial role in framing our
knowledge and constructing certain orders” [12]. Fig. 2 effectively demonstrates the scope of modern economic and socio-cultural activities where the Internet, web and communication are employed [10].

Fig. 2: Serelex Semantic Proximity Graph for the lexemes “Communication”, “Internet”, “Web” [13].

It is taken for granted that electronic media and the Internet are key players in the field of communication. The central idea of my hypothesis is that besides all digital devices and gadgets one should not diminish the power of words in modern global socio-cultural digital community. As S. Benhabib has masterfully formulated: ‘To be and to become a self is to insert oneself into webs of interlocution’ [3]. But which linguistic items are of primary importance for digital communication? My answer is the domain of idioms or set communicative formulae, as they represent culture connotations, codes and symbols, describing the ideas of ‘good’ and ‘bad’ in every society [15;16]. Even modern programming languages have developed their special type of idioms – programming idioms.

Fig. 3: Programming Idiom Graph [2].

Idioms are ubiquitous in the history of mankind. In his spectacular work Jeffrey C. Alexander specified the road, taken by the notion of Holocaust from the simple idiom of ‘atrocities of war’ to the Testament idiom ‘the dominant evil of our time’ [1]. The capacities of digital space to use idioms for image creation are even greater. Here the question arises: do digital media professionals completely fulfill the potential of idioms in electronic communication? Do people realize the importance of idioms for their socio-cultural interaction? In order to answer these questions I’ve conducted the following analysis.

Data and Methods:

Stage 1:

In order to prove the fact that people use idioms in their day-to-day communication, which is a model for the electronic Internet communication, I conducted an omnibus survey in cooperation with ‘Yuri Levada Analytical Center’ [17]. Levada-center carries out nationwide polls in Russian Federation, which are based on a representative sample of 1600 people over 18+ from 130 sampling points across 45 regions of Russian Federation.

Our sample did not include the Chechen and Ingush Republics, as well as remote and sparsely populated region of the Far North (Nenets, Yamal-Nenets, Taimyr Autonomous Region, the Evenki Autonomous District, Kamchatka, Chukotka, Sakhalin Oblast). There were 9 Russian areas, totally excluded from consideration. The adult population in the excluded areas did not exceed 4 %. Residents of small settlements with a population of less than 50 people (about 0.8 % of the adult population of Russia), the military (about 0.8 % of the adult population of Russia), the person in custody or detention during the investigation (about 0.8 % of the adult
population of Russia), as well as homeless people (up to 1.5 % of the adult population of Russia) were not included to the survey.

In each federal district independently all the primary sampling units were divided into strata according to their population number:
1. The city of more than 1 million people;
2. The town numbering from 500 thousand to 1 million people;
3. The town numbering from 100 thousand to 500 thousand people;
4. Urban settlements of up to 100 thousand people;
5. Villages.

These 5 categories were further subdivided into smaller groups. Given the characteristics of population distribution in the regions, we formed 36 strata. The total sample (1,600 respondents) was distributed proportionally among all strata of the adult population in each stratum. The number of primary sampling units was determined from the restrictions on the average number of respondents in the same town / rural area (7-13 respondents).

Our central survey question was as follows: ‘Do you use common routine expressions, quotations from films and public speeches of famous people, bad language set expressions in your day-to-day life”? All these linguistic terms are referred to the domain of idioms in Russian.

The results show that male respondents are more likely to admit the fact of idioms usage in their communication, than female respondents (43.4 % against 42.8 % respectively). This idea has a very important practical implication. Nearly half of all the respondents admitted their usage of idioms in day-to-day communication. Thus, idioms have to be used in electronic media communication as well.

Stage 2:

Word combinations and idioms are analyzed by checking their frequencies using the following on-line resources: BNC [4] and COCA [6]. This type of analysis is very important as it helps demonstrate to which variety of modern English, British or American, this or that idiom belongs. ‘The Corpus of Contemporary American English (450 million words) is more than four times as large as the British National Corpus (100 million words). As a result, it often provides data for lower-frequency constructions that are not available from the BNC. … COCA and the BNC complement each other nicely, and they are only large, well-balanced corpora of English that are publicly-available. The BNC has better coverage of informal, everyday conversation, while COCA is much larger and more recent, which has important implications for the quantity and quality of the data overall’ (Corpus of Contemporary American).

For frequencies analysis I’ve chosen word combinations with the following components: body, head, and heart components. Consider the following collocations with high frequency indexes. A human body = 2207 (BNC and COCA combined), dead body = 1343 (BNC and COCA combined), whole body = 1217 (BNC and COCA combined), upper body = 1070 (BNC and COCA combined), entire body = 733 (BNC and COCA combined), female body = 411 (BNC and COCA combined), male body = 139 (BNC and COCA combined).


Idioms with the highest frequencies:


The analysis demonstrates an American tendency to associate body with a tangible economic result: governing body, student body. Body is used in arts, show business and sports to generate profit and attract attention. ‘Head’ component is also very frequent in modern communication. To bang one's head against a brick wall – to be doggedly attempting the impossible and suffering in the process [BNC – 1; COCA – 7 (2002-1990)].

To be hanging over someone's head (of something unpleasant) – to threaten to affect someone at any moment [BNC – 7; COCA – 40 (2012-1990)]. To be on someone's head – to be on someone's sole responsibility [BNC – 2; COCA – 4 (2012-1990)]. To bite someone's head off – to reply sharply and brusquely to someone [BNC – 4; COCA – 17 (2012-1990)].

To hold up one's head – be confident or unashamed [BNC – 5; COCA – 26 (2012-1990)]. To keep one's head above water – avoid succumbing to difficulties, typically debt [BNC – 6; COCA – 16 (2012-1990)]. To make head or tail of – [usu. with negative] understand at all [BNC – 13; COCA – 19 (2012-1990)]. To take it into one's head to do something – to imputiously decide to do something [BNC – 16; COCA – 24 (2012-1990)].

Idioms with the highest frequencies:

To put their/our/your heads together – to consult and work together [BNC – 19; COCA –99 (2012-1990)].


To come to a head – to reach a crisis [BNC – 149; COCA – 363 (2012-1990)].


From the bottom of one's heart – to speak honestly. [BNC – 15; COCA – 137 (2012-1990)].


By heart – from memory. [BNC – 173; COCA – 675 (2012-1990)].

Idioms with the highest frequencies:


To break smb.'s heart – to charm, to fascinate [BNC – 92; COCA – 919 (2012-1990)].

The analysis presented shows it very clearly that idioms are frequent in day-to-day communication. Thus, the rules of idiomatic usage are important to our understanding of communicative process in real life and the Internet. They serve as socio-cultural indicators of speakers’ preferences and must be implemented for information retrieval.

Discussion:

My research demonstrates that there are various theoretical and practical implications of my analysis. My survey is limited to English idioms, though now I proceed to the stage, where I discuss features common to idiomatic systems of many languages.

As any information system, the system of social communicative signs of idioms can be described in terms of mathematical laws. My central idea is that idioms should be analyzed within the framework of information maximum. The principle of information maximum firstly appeared in the works by Shannon [14] and later it was developed by Soviet and Russian scholars [7]: ‘the system tries to find such response y, which enables the maximum of useful information about the given stimuli x’.

Natalya Zavyalova,2014

L (X,Y) = H (Y) – H (Y/X) – β R (X,Y) → max,
where H (Y) – the entropy of the system’s conditions, H (Y/X) – the entropy of the system’s mistakes and its reactions to outer stimuli, R (X,Y) – the average consumption of the system’s resource in Y condition within all X environmental conditions, and β – the indicator of the system’s deficit (β=0, when the system has an unlimited energy resource, n β=1 when the system has a very big energy deficit)” (Ibid.).

According to the described above principle, all informational systems develop within 3 main strategies:

Expansion is realized through new varieties and numbers of environmental conditions, in which the analyzed system can exist. This tendency is sometimes called ‘searching behavior’ tendency H (Y). While idealization is concentrated on the tendency to minimize the system’s mistakes entropy, idealization is sometimes known as the ‘conservative behavior’ tendency H (Y/X). It is achieved mainly through the repetition of the best variants. Resource saving is realized in 2 possible ways: 1) minimizing the resource spent R (X,Y); 2) minimizing the indicator of the system’s deficit β by means of resource growth. My research is limited to the analysis of British and American idioms, their strategies of idealization, reflected in the most frequent idioms.

As in other languages expansion of British and American idiomatic system is realized by means of describing various fragments of day-to-day discourse in idioms. One and the same object of day-to-day routine may be used in various contexts. Idealization strategy of British and American idioms is analyzed by checking frequencies using BNC (British National Corpora) and COCA (Corpora of Contemporary American). Resource saving of British-American idiomatic system is achieved by the following strategies: 1) alliteration, assonance; 2) 2-word idioms; 3) parallel constructions; 4) repetitions; 5) polysemy and synonymy; 6) decomposition of proverbs and sayings; 7) borrowings from other languages and literary works of English-writing authors; 8) simple verb forms in habitual aspect, rather than verb forms in continuous or perfect aspects, which require an auxiliary verb and a verb. This finding leads to the conclusion that electronic communication is driven by social stereotypes and codes, which serve as indicators for information retrieval.

Summary:
The practical implications of my study lie within the domain of social influence in electronic communication area, which is driven by communicative idiomatic stereotypes [5]. The theoretical value of this research is the expansion of idiom status, whose role for the electronic communication analysis has been minimized to purely linguistic phenomena. I proceed from the idea that idioms are social in nature. Thus, their careful analysis adds to the profound understanding of social life and economic environment, as A. Samuels rightfully put it ‘disputes concerning human nature underlie many debates on economic theory’ [11]. I suggest the idea, that communicative formulae and idioms should be more widely accepted for information retrieval, as they perform the function of national preferences’ indicators. The results of these findings are relevant to multicultural societies, migration adaptation practices and global business development. Further research should contribute to analysis of expansion and resource saving strategies, frequency indexes of other idiom clusters and other languages.

REFERENCES


