Intelligent Dictionary Interfaces: Usability Evaluation of Access-Supporting Enhancements

Anna Sinopalnikova and Pavel Smrž

Faculty of Information Technology, Brno University of Technology
Božetechova 2, Brno 61266, Czech Republic
{sino, smrz@fit.vutbr.cz}

Abstract

The present paper describes psycholinguistic experiments aimed at exploring the way people behave while accessing electronic dictionaries. In our work we focused on the access by meaning that, in comparison with the access by form, is currently less studied and very seldom implemented in modern dictionary interfaces. Thus, the goal of our experiments was to explore dictionary users' requirements and to study what services an intelligent dictionary interface should be able to supply to help solving access by meaning problems. We tested several access-supporting enhancements of electronic dictionaries based on various language resources (corpora, wordnets, word association norms and explanatory dictionaries). Experiments were carried out with native speakers of three European languages – English, Czech and Russian. Results for monolingual and bilingual cases are presented.

1. Introduction

Becoming extremely rich in content, modern electronic dictionaries remain parsimonious in supporting full-featured access to this content. Their interfaces are usually able to supply only access by form services, such as misspelling correction, or less often implemented homophony checking (Zock & Fournier, 2001; Bilac et al., 2003; OneLook), or morphology analysis (Breidt & Feldweg, 1997, Maxwell & Poser 2004). The access by meaning is rarely implemented. The main ideas in this area are presented by thesaurus construction (Roget, 1852; Miller, 1985) supposing a development of a service "that allows user to explore an on-line dictionary on the basis of semantic, rather than alphabetic, similarities" (Fellbaum, 1998), lookup on the basis of associations (Zock & Bilac, 2004), and access by reverse dictionary (Bernstein, 1975; Glanze, 1990, Edmonds, 1999) that in practice resulted in the definition search (AHDAL; CALD; CCED; OneLook).

Access by meaning is not yet studied and elaborated enough. Consequently, services supplied by available dictionary interfaces often provide inappropriate or insufficient output, which only confirms the need for intelligent, structured and differentiated approach to those problems. The necessity to obtain empirical data and test evidences was the primary motivation of the work presented in this paper.

To obtain evidences on human behaviour in access problem situations was the main goal of our work. We believe that studying dictionary users' needs can provide a necessary empirical base for designing an effective intelligent dictionary interface. Our secondary goals were as follows:

- To prove the hypothesis that different types of words cause different types of access problems, and different problems necessitate different solutions, in particular, need different types of access-assisting resources to be applied.
- To explore in what way language resources (LRs) can facilitate the lexical access and, what strategies of access they can support and in which cases.

- To evaluate and compare the effectiveness of LR-based enhancements in solving access by meaning problems.

2. Design of experiments

It is not a trivial task to devise a procedure for testing word-access strategies. In our experiments, we decided to simulate various situations where people experience word access problems and would need an assistance of a dictionary. These situations could be roughly divided into: "Tip-of-the-Tongue" (TOT)[1] cases when people could not remember a word they are sure they know, and "Don't Know" cases when people search for an unknown word. There is no strict border between these two – sometimes it is difficult to distinguish the two and there is also a variety of intermediate cases. What concerns the purposes of our work, Don't Know cases turned to be as informative as TOTs[2]. Thus, in the following passages, we refer to them jointly as word access problems, still having in mind the existing differences.

2.1. Subjects

Experiments described below concern inducing different lexical access problems, and were designed to explore human behaviour while solving such problems.

Subjects that participated our experiments were between the ages of 25 and 78, both male and female, of higher education and various professional background (administration, chemistry, computer science, linguistics, mathematics, mining and public relations). Subjects were

---

[1] For the psychological surveys on this phenomenon see (James 1890, Brown and McNeill 1966, Maylor 1990, Burke et al 1991).

[2] Roughly speaking, the main difference between TOT and Don't Know cases concerns the fact that experiencing the TOT state, people are able to describe (approximately) not only the meaning of the word, but also its form. But as we are interested in studying access by meaning problems, both states are equally informative. Even not knowing a word, people can describe its meaning through the definition-like approximation, or produce synonyms, hypernyms , antonyms of the word, and they know in what context or situation they want to use the searched word in.

native speakers of Russian and Czech. The former participated in both monolingual and bilingual access experiments; the latter took part only in bilingual settings. What concerns bilingual settings, all Czech subjects knew English and Russian at advanced level; all Russian subjects knew English at advanced level, and were able to communicate and produce technical texts in English.

2.2. The experimental procedure

The general experimental procedure was divided into two steps: i) collecting data on the word access problems and ii) processing the data using various LRs.

The first step was to induce lexical access problems, to identify and study the TOT and Don't Know cases, to record the way the subject proposes to access the word (i.e. word approximations) and to store this information together with the correct words the subjects had in mind. The second step was to process the results, simulating an intelligent search assistant that would select the optimal access-supporting method every time and to compare the results with the 'blind' case

Collecting the data, we simulated situations inducing word access problems of the following types:
1. In the **first** series of experiments, twelve subjects were asked questions that could be correctly answered by a single word, e.g. *"What word means to formally renounce a throne?"* – with the expected answer *to abdicate.* In case a subject reported TOT or Don't Know state, potential input to the ideal dictionary interface was recorded – i.e. all that subject knew about the word (similar words, appropriate contexts or situations, its domain etc.) and the way the subject proposed to look for the word he/she needed.
2. The **second** series of experiments included a variant of the first series. The subjects were asked to fill the gap in the sentences like *"The colour of the clear sky is blue, but the colour of the deep sea is___"* with the possible answers *ultramarine, indigo, azure, cerulean, navy* etc.
3. The same subjects participated in another variant of the experiment (the **third** series), when sentences were replaced with pictures. The subjects were asked to describe an object or a situation presented.
4. In the **fourth** series of experiments, four subjects were asked to log all word access problems they experienced while writing a paper. Their 'diaries' containing information about the starting point of the searches – what subjects remembered about the word (meaning, synonyms, context, domain etc.) and the way they found the correct word – were then processed.

The experiments were organised around the low-frequent words to guarantee as much word access problems as possible. We conducted experiments with Russian and English nouns, verbs, and adjectives.

The experiments were carried out in monolingual settings (Russian-Russian), and then repeated in bilingual ones (Russian-English, Czech-English, Czech-Russian). In the case of access problems in a foreign language, the subjects might approximate the searched word using foreign language as well as their mother tongue.

Apart from the fourth series, the described experiment settings correspond to the search with no feedback.

Subjects faced the instructor, not the dictionary interface itself. They got no information whether the access method they chose failed or succeeded. They have no chance to correct their input according to the dictionary output and repeat their search.

3. Results

3.1. Word approximations

The experiments described in the previous section produced quite heterogeneous data. The experiments differ in the naturalness of the situations simulated and processability of the results. For instance, being the least natural situation, question-answering supplies the most easily structured and processable results, while the writing-a-paper settings deal with the real (most natural) word access problems but produce the most serious obstacles for identification and extraction of data we are interested in.

A unified form of the data representation was called **word approximations (WAs)** – unstructured descriptions of the words given by the subjects in TOT or Don't Know situations. To convert the answers to WAs, data were reduced to the access by meaning constituents, thus excluding all access-by-form descriptions, e.g. approximation of *ignorant* through *sounds like arrogant,* as well as non-verbal descriptions (gestures, mimics, sounds etc.). As WAs were to be unstructured lists of separate meaningful words (sometimes, phrases) for the blind approach (see below), we also excluded all the functional constituents, e.g. *do, have, no, and, something, like,* and performed the simplest morphological analysis, e.g. transforming *carrying* to *carry*, *players* to *player* etc. Thus, for example, we obtained the following word approximations:

Orphan – who has no mother, no father, step-mother and step-father, alone, poor, crying. → *mother, father, step-mother, step-father, alone, poor, cry.*

Superficial – sounds like "surfing", similar to "top", occurs with "polish". → *top, polish.*

Cuddy – a man carrying a bag for the player, criket-assistant. → *man, carry, bag, player, cricket, assistant.*

The main body of the word approximations was obtained in the first series of experiments. Other series produced a significant supplement to it. Conducting experiments with 900 questions for foreign languages and 800 for the mother tongue, we collected 79 and 34 WAs. After processing the diaries, we obtained another 12 WAs for mother tongues and 25 for foreign languages. Together with 33 and 59 from the filling-gap and describing-picture experiments, we end up with 89 word approximations for mother tongue and 163 for foreign languages.

3.2. Processing the results

The complete set of word approximations was processed using LRs of four different types (corpora, wordnets, word association norms and explanatory dictionaries) and two approaches of the word access (blind and intelligent one). The following English and Russian LRs were applied to the data processing:

1. **Wordnets (WN):**
   - Princeton WordNet 2.0 (115 000 synsets);
   - RussNet 0.2 (5500 synsets).

Dictionary-search method	Number of "winners"			
	for the blind approach		for the intelligent approach	
	mother-tongue	foreign language	mother-tongue	foreign language
Fulltext search in entries	10	12	11	15
Corpus-based collocates	8	5	7	2
Word sketches	N/A	14	N/A	17
WN-based interface	1	15	1	17
WAN -based interface	9	9	9	6

Table 1: Results of experiments showing how many times each particular word access method was the most efficient

2. **Word Association Norms (WAN)**:
   - EAT – Edinburgh Word Association Thesaurus by Kiss et al (1972) (54000 words covered – 1000 subjects);
   - RAT – Russian Word Association Thesaurus by Karaulov et al (1994–1998): (23000 words covered – 1000 subjects).
3. **Text corpora:**
   - BNC – British National Corpus (112 mln);
   - Bokrjonok 4.0. – balanced corpus for Russian (21 mln).
4. **Word Sketches:**
   - WordSketches for English collected from BNC (Kilgarriff et al., 2004); not available for Russian.
5. **Dictionaries:**
   - LDOCE – Longman Dictionary of Contemporary English;
   - EDR Explanatory Dictionary of Russian by Shvedova and Ozhegov (1992);
   - MAD – Multitran dictionaries (used for bilingual search).

The simplest way to process the data (called **blind approach**) was to take the logged approximations as sequences of words with no structure or special relations. These undifferentiated word approximations were then used as the input for each of the access-assisting resource. That means that even if a subject claimed that a word X is, e.g., a more general term (hypernym) of the searched word Y, this information was not used to focus the search in the blind approach. On the other hand, using the information from subjects and performing a detailed (semantic) analysis of the WAs, we were able to differentiate particular types of relations. In the second case, the WAs were structured and processed in the way optimal for the given method (**intelligent approach**).

Five word-access methods (with four respective language resources) were evaluated against both blind and intelligent settings. We entered the word approximations as an input to all the access-assisting resources and analysed their outputs. Regularly, the correct word was returned by more than one access-assisting LR. We computed the position of the searched word in the output. The reported 'winner' is the method that returned the correct word at the upper position. The results of the experiments are summarized in Table 1.

4. Interpreting the results

4.1. Effectiveness of LR-assistance in solving word access problems

The presented numbers were obtained for quite a small group of subjects. Thus, they are not perfectly suitable for generalisation and drawing accurate conclusions. Rather they hint possible directions of the following considerations.

Comparing the numbers for different access-assisting enhancements, one can see that in the **bilingual** settings **wordnets** perform good and that their usage for the dictionary look-up is comparably effective to a sophisticated corpus-based LR, namely Word Sketches. But the effectiveness of wordnets in the **monolingual** (Russian-Russian) settings is surprisingly low. The possible explanation lays in the difference between English and Russian WNs in nature and size. Princeton WordNet is a large database, covering not only the basic vocabulary but also several terminological domains. While RussNet, covering only the core of the Russian language, presents only the basic vocabulary that hardly evokes TOTs or Don't Know states for native speakers.

The relatively high efficiency of **WAN** in assisting word access is comparable to an extremely simple corpus-based method of WA processing. This could be explained by the very nature of the resource, which combines the features of the semantic network with that of the text corpora. Extremely important for its effectiveness turned to be the exposure of the domain relations, e.g., between *needle* and *thread,* or *reflex* and *knee.* This knowledge was regularly used by the subjects as a basis of the WAs and is otherwise accessible only with corpus-driven methods.

The high numbers obtained for the full-text search in **dictionary** entries were caused mostly by the inclusion examples into the entries, and partly by the disproportion of the WA types processed (see Section 4.2.1). The main part of WAs collected in our experiments was covered by nouns, and for them the search in definitions is quite effective, which does not always holds for verbs and adjectives.

POS of the searched words	Dictionary-access method				
	Full text search in definitions	*Corpus-based collocates*	*Word sketches*	*Wordnet-based interface*	*WAN -based interface*
Nouns	14%	12%	32%	29%	13%
Verbs	8%	25%	26%	25%	35%
Adjectives	19%	24%	17%	18%	20%

Table 2: POS-preferences for different dictionary-access methods (intelligent bilingual settings)

The results obtained for the **corpus**-based methods of the dictionary access (simple and sophisticated ones) provide evidence for the effectiveness of applying such LRs for interface enhancements. Our findings confirm the earlier ideas by Zock and Bilac (2004) The N/A results mean that the given LR, namely the **WordSketches** for Russian, are not available at the moment.

The smaller numbers in the column of the intelligent approach in comparison with the blind approach for corpus-based and WAN-based interfaces should not be interpreted wrongly. They do not indicate that the blind approach outperforms the intelligent one in these cases. The results only suggest that the detailed analysis of WAs (user's input) contributes to the effectiveness of other LRs too, and in larger extent.

4.2. Factors that affect the choice of word access strategy

In our experiments, we have observed several factors that affected the way people approximate the searched words. They are infinite in number, but among the most important we should mention the following:
1. The most obvious factor affecting the choice of access strategy is the characteristics of a **subject**/dictionary user (e.g. age, level of education, cultural background, occupation etc.). It influences the frequency and type of access problems, the way people recall and approximate words, the choice of the assisting resources etc. For example, older subjects reported more access problems. Mathematicians tended to classify and structure their answers even without instructions given to them e.g. *superficial – sounds like "surfing", similar to "top", occurs with "polish"* Computer scientists had difficulties with accessing the word *violet,* but not *magenta.* These factors are hard to be reflected in the dictionary interface in any near future.
2. Naturally, the type of access problems – **monolingual or bilingual** – has also an impact on the way people initiate the search of the word When accessing a **bilingual** or **multilingual** dictionary, the user tend to limit access procedure to an equivalent in the second (other) language, rather than a description in the same language. In case of **lexical gaps** a user applied more ways of word approximation, e.g. corpus-, WN- or WAN-based ones using his mother-tongue or a target language.
3. The **availability** (and the quality/size) of the resources can play the most important role in the choice of the access methods.
4. The access process obviously differs in connection with the **media** and devices we use to access an electronic dictionary.

5. The most interesting were **linguistic characteristics** of the word that influenced the type of access problems and the process of access strategy (e.g., frequency of the word, familiarity, longitude, POS, proper vs. common, concrete vs. abstract, number of synonyms, number of senses).

The non-linguistic factors were discussed in (Sinopalnikova & Smrz, 2006), here we will focus on the linguistic ones. In general, we have proved our starting hypothesis that (linguistically) different types of words cause different types of access problems and necessitate different solutions and, thus, need different access-assisting LRs to be applied. In the following subsections we discuss our findings on two such factors, namely how Part-of-Speech (POS) and abstractness/concreteness of the words correspond to the type of access problems they evoke and the optimal strategy used to access them.

4.2.1. POS factor

Firstly, we have found that words of different POS evoke access problems with different frequency. For example, verbs and adjectives caused TOT and Don't Know states less often than nouns. Nouns covered about 68% of access problems. WAs of verbs and adjectives were collected mostly in the first experimental settings – that of answering-questions. This could be explained by difference in the nature of words – by the fact that names of qualities and relations are not so numerous as names of objects, and subjects, if not pushed, easily avoid or replace unknown verb or adjective with known one.

Secondly, we have observed that word approximations of words of different POS tend to be effectively searched with different LRs. The efficiency of access-assisting LRs and respective methods is distributed as presented in Table 2.

Nouns were regularly approximated by a (quasi)definition containing a natural genus proximum and specific features, or synonyms, co-hyponyms, or hyponyms, e.g. *instrument to measure a blood pressure (manometer); monk, Carmelite (nun)*. Thus, it was only natural that search in definitions and assistance by WNs or WordSketches, which have a thesaurus options, had a better return in accessing nouns. Words with not so clear signification, such as verbs and adjectives, were rarely described through definitions. Rather they tended to be approximated by semantically related words, domain relations and typical collocates, e.g. *cause, reason, consequence (necessary), vital, crucial (necessary)*. Still for adjectives we have not observed any clear preferences in the access-assisting strategies.

4.2.2. Concrete vs. abstract words

The difference between **concrete** and **abstract** entities affects the way people describe the meaning of a word. It

was rather interesting observation that for both concrete and abstract words the subjects gave WAs that mainly consist of concrete words. E.g. *reflex* was approximated by *jerk, knee* that evidently refer to the typical example of the reflex test, when a tapping below the knee causes the leg to suddenly jerk forward. Obviously, that evidences for the difference in relations between search words and the words presented in respective WAs, thus explains the difference in the effectiveness of access strategies and respective LRs:

- Concrete words – WNs;
- Abstract words – WAN, corpora.

5. Conclusions

Our experiments were focused on the access by meaning, not by form. We tried to explore what language resources can facilitate the lexical access and in what way, what strategies of access are preferential and in what cases. In our experiments, we simulated monolingual look-up (Russian, Czech) as well as bilingual one (Russian-English, Czech-Russian, and Czech-English).

The experiments were conducted with the low frequent nouns, verbs and adjectives. For each word, we calculated the correspondence between access-affecting factors (linguistic ones) and the access strategies that gave the best results in the blind and intelligent search. These data may serve as necessary empirical ground for designing an effective intelligent dictionary interface capable of optimal choice of the access-assisting strategy and of appropriate structuring of the dictionary search output.

The described experiment settings correspond to the simplest search only. Our future research will focus on advanced techniques of the access evaluation. We will provide the user with a more realistic application interface, which will be able to present search results from more than one information source simultaneously and allow the user to add new data to its current input and see the updated results. We will also pay attention to the design of the search form. The current prototype is rather complex. We cannot expect that an average user will be able to classify his/her word approximations correctly, to define the kind of relation between given word and the searched one or to identify optimal access strategy.

6. Acknowledgements

7. References

7.1. Dictionaries

AHDAL – *The American Heritage Dictionary of the English Language:* Fourth Edition. 2000. http://www.bartleby.com/61/

Bernstein, T. (1975). Bernstein's *Reverse Dictionary*. NY:Times Books.

CALD – *Cambridge Dictionaries Online – Cambridge Advanced Learner's Dictionary* http://dictionary.cambridge.org/

CCED – *Collins COBUILD Advanced Learner's English Dictionary*. 4th edition. Harper Collins Publishers, 2003.

COED – *The Compact Oxford English Dictionary* – http://www.askoxford.com/

Edmonds D. (1999) *The Oxford reverse dictionary*. NY: Oxford University Press.

EDR – Shvedova N. V., Ozhegov S. Ju. (1992) *Explanatory Dictionary of Russian*. http://www.lib.ru/ozhegow.

Glanze W.D. (1990) *Illustrated Reverse Dictionary: Find the Words on the Tip of your Tongue*. Pleasantville, NY.

LDOCE – *Longman Dictionary of Contemporary English* (Online) http://www.ldoceonline.com/

MAD – *Multitran automatic dictionaries* (Online). http://www.multitran.ru

MW – *Merriam-Webster Online Dictionary* http://www.m-w.com/

OneLook – http://www.onelook.com/

7.2. Bibliography

Bilac S., Baldwin T., Tanaka H. (2003). Improving dictionary accessibility by maximizing use of available knowledge. *Traitement automatique des langues*, *44 (2)*.

Bilac, S.,Watanabe, W., Hashimoto, T., Tokunaga, T. and Tanaka H. (2004). Dictionary search based on the target word description. In *Proceedings of the 10th Annual Meeting of the Association for Natural Language Processing.* pp.556–559.

Breidt, E., Feldweg, H. (1997). Accessing Foreign Languages with COMPASS. *Machine Translation Journal, Special issue on New Tools for Human Translators, 12.* pp. 153–174.

Brown, D.M., McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5. pp. 325–337.

Burke, D.M., MacKay, D.G., Worthley, J.S., Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults. *Journal of Memory and Language, 30*. pp. 542–579.

El-Kahlout, I. D., Oflazer, K. (2004). Use of Wordnet for retrieving words from their meanings. In *Proceedings of the Global Wordnet Conference (GWC2004).* pp. 118–123.

Fellbaum, C. (ed.) (1998). *WordNet: an electronic lexical database*. MIT Press.

James, W. (1890). *The principles of psychology*. Vol. 1. New York: Henry Holt & Co.

Karaulov, Ju.N., Cherkasova, G. A., Ufimtseva, N.V., Sorokin, Ju. A., Tarasov, E.F. (1994, 1996, 1998). *Russian Associative Thesaurus*. Moscow.

Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D. (2004) The Sketch Engine. In *Proceedings of the 11th European Association for Lexicography) International Congress (EURALEX2004),* Lorient, France, Université de Bretagne Sud. Volume I. pp. 105–116.

Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In: Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (eds.) *The Computer and Literary Studies.* Edinburgh: University Press.

Löfberg, L, Juntunen, J-P, Nykanen, A, Varantola, K, Rayson, P, Archer, D. (2004). Using a semantic tagger as dictionary search tool. In Williams G. and Vessier S. (eds.) *Proceedings of the 11th EURALEX (European Association for Lexicography) International Congress (Euralex 2004),* Lorient, France, 6–10 July 2004. Université de Bretagne Sud. Volume I, pp. 127–134.

Maxwell M., Poser W. (2004). Morphological Interfaces to Dictionaries. In *Proceedings of the Workshop on Enhancing and using electronic dictionaries, COLING 2004.*

Maylor, E.A. (1990). Age, blocking and the tip of the tongue state. *British Journal of Psychology*, *81*. pp.123–134.

Rastle, K. G, Burke, D. M (1996). Priming the tip of the tongue: Effects of prior processing on word retrieval. in young and older adults. *Journal of Memory and Language, 35*. pp. 586–605.

Roget, P. M. (1952). *Everyman's thesaurus of English words and phrases.* London, New York, Dutton.

Rychly, P., Smrz, P. (2004). Manatee, Bonito and Word Sketches for Czech. In *Proceedings of the Corpora–2004. Saint-Petersburg,* Russia. pp. 34–41.

Sinopalnikova, A., Smrz, P. (2006) Knowing a Word vs. Accessing a Word: Wordnet and Word Association Norms as Interfaces to Electronic Dictionaries. In *Proceedings of the 3rd International Global WordNet Conference (GWC-2006).* Jeju, South Korea. pp. 265–272.

Zock, M., Bilac, S. (2004). Word lookup on the basis of associations: from an idea to a roadmap. In *Proceedings of the Workshop on Enhancing and using electronic dictionaries, COLING 2004*. pp. 29–35.

Zock, M., Fournier, J.-P. (2001). How can computers help the writer/speaker experiencing the Tip-of-the-Tongue Problem? In *Proceedings of the RANLP-2001,* Tzigov Chark. pp. 300–302.