

# Lexaurus: A Multidimensional, Ontology-based Bilingual Electronic Dictionary\*

Jonathan C. Oh and Hyo-Pil Shin

## 1. Ontology, Dictionary and Thesaurus

Word lists play an indispensable role in natural language-based communication. Lexical items in a word list can be ordered alphabetically as in dictionaries or arranged in categories by their meanings as in thesauri. Dictionaries currently available are basically one dimensional, words with their phonetic, syntactic, and semantic characteristics are listed in a one dimensional sequence, an alphabetic ordering. Relationships among words in essence are not represented except in occasional synonym or antonym lists. In a monolingual dictionary, semantics is often expressed in terms of synonyms. In both monolingual and bilingual dictionaries, the treatment of semantics is not principled. No two dictionaries list the semantic items in the same order. In fact, it is impossible to see any system in the semantic structure in a word entry. Just pick up any dictionary, select any common entry, and try to figure out the underlying semantic structuring. If one feels that he has found some semblance of a semantic theory underlying the arrangement of meaning items in an entry, further explore to see if the same theory is used in any consistency. Now pick up any pair of more authentic dictionaries and duplicate the above outlined experiment. The result is about the same. The same chaos is found in bilingual dictionaries as well.

In a thesaurus, words are listed according to their semantic categories. An ontology is represented in this categorial arrangement. Relationships among words albeit rudimentary ones are expressed. A thesaurus, for Roget's, can be very useful especially if equipped with a comprehensive

\*This research is supported by a grant from Korean Engineering Center, Daejeon, Korea.

index. For those aspiring to use languages forcefully and precisely, a thesaurus can prove to be a very necessary tool. What we are proposing here in this paper is somewhat akin to thesauri, except the fact that thesauri are monolingual in nature. The services thesauri offer to monolingual communities need to be provided for bilingual purposes as well. There is nothing comparable to monolingual thesauri in the bilingual field. The only kind of word lists available in a bilingual field is dictionaries. Dictionaries aim to provide only basic meanings, rarely are ambitious enough to make available a rich lexical resource from which to choose a pragmatically, semantically, and stylistically adequate equivalent expression in the target language. Translators are offered a basic set of meanings for a given entry and left in the lurch in finding an equivalent. A creative processing as in creative writing is then called for.

The sort of restrictions mentioned in the previous two paragraphs was largely a result of constraints on publication means. With a two dimensional printed page representation, lexical relations were virtually impossible to capture. With the electronic computational resources now available, we can hope to provide multidimensional services. Colonies of semantically related words should be extractable, ordered along many frequently used parameters such as degrees of intensity, of formality, of technicality, of sincerity/jocularly, of familiarity, of generality, of vernacularity, of permanency, and of lexicalization, speech level, speaker's attitude toward the concept expressed, the speaker-hearer relation, discipline/specialty/area of use, and the gender of the speaker. The underlying data structures are not a sequence or a list, not even a tree, but a general graph or network, a multidimensional one at that.

A natural language may be viewed as a set of specialized languages constantly under change. One can appreciate the truth of this statement by considering the ever-changing legal, medical, and computer jargons. In actuality, such diversity of language is much more pervasive and intrinsic than usually recognized. It is not inconceivable that in near future lexical bases of various kinds are maintained at different locations and institutions so that they can be made available through networks. These distributed multidimensional lexicons with the kind of colonial extraction function mentioned above should offer a new generation of lexical information services.

## 2. Lexaurus, A Multi-dimensional, Ontology-based Bilingual Dictionary

In this paper, we would like to focus on Lexaurus, a multidimensional ontology-based bilingual dictionary that is designed to overcome the deficiencies discussed in the previous section. Lexaurus is an electronic dictionary whose entries are alphabetically ordered, and the target language equivalents are listed under each entry according to the basic ontology that is adopted. We will try to enlist these equivalents as exhaustively as possible. Each equivalent is an element of several semantic colonies. A semantic colony is an organized group of categorially related words/expressions along one of the parameters mentioned earlier. So while only a representative is listed under a given entry, if one wishes one could open up the whole colony the word represents and further explore for a better match.

In order to see the necessity of multidimensionality, let us consider the task of a translator. We live in the time of information exchange between cultures in explosive demand. Translation need is greater than ever before. Demand for quality translation has also dramatically increased. Anyone who has ever been involved in any practical translation project painfully realizes that a realistic dictionary that meets everyday demands by enlisting proper equivalents is not currently available. Most of currently available dictionaries are semantically oriented in the sense that word choice is not given due consideration but that they attempt to capture only the basic meaning of words. In a quality translation, more than just the guide to the basic meaning is needed.

A translator given a source expression searches for as close an equivalent as possible, a proper target expression that matches the source text as closely as possible not only in the basic meaning but in various stylistic and pragmatic aspects. The first step here is to characterize such equivalency. He needs to find equivalent expressions to the source text expressions which are lexically, syntactically, semantically, and pragmatically adequate. The first approximation is the semantic equivalence, and the source expression and the target expression must be semantically equivalent. The semantic content of an expression we are looking for is a function from the use context to the denotation, not just the denotation per se. A rather dramatic example illustrating the point is the kinship term. One of the authors was interpreting a talk by an English-speaking speaker into Korean. The story

started out introducing a man and his brother-in-law. The translator was simply at a loss which Korean kinship word to use without the benefit of knowing whether the in-law is through the man's brother or sister and the age relation between the man and his sibling involved here.

This means that the assumption behind the current practice in lexicology is mistaken. It appears to be generally assumed that there is a typical context, a sort of standard one for which a dictionary may be constructed nonfunctionally. The importance of isolating parameters that are involved in defining a context of use cannot be overstressed. In order to maintain such semantic equivalence a dictionary must organize all possible equivalents according to these parameters so that when an expression is given to a translator he will apply the semantic function represented in a dictionary to the context he could best identify to select a hopefully small set of equivalents.

Multidimensional bilingual dictionaries can be somewhat standardized by being based on ontology. Colonies need to be recognized and organized according to the adopted ontology. The equivalent lists under each dictionary entry need to be organized according to the ontology.

### 3. Lexaurus and Thesaurus

We have been working on Roget style thesaurus for the Korean language. Initially we started out to basically translate Roget's Original Thesaurus. We have quickly decided that Roget's underlying ontology is not transferable to Korean. Furthermore, it has become obvious that there is not one ideal ontology that can be used as a standard one<sup>1</sup> Also for a given ontology, there may be many views. Grocery stores, supermarkets, and drug stores, for instance, have developed their organization of sales items. There are many considerations that affect the organization. It is not just to match customers' view although that is the primary factor to keep in mind. After all, the profit margin depends on the volume of sale and the reduction of extra service needed which in turn is critically related to the degree of ease with which customers find their desired items. Other factors that influ-

<sup>1</sup> We are not making any theoretical claim here concerning whether there can be such a universal ontology. We are merely saying that no such ontology is currently available or is expected to be in any foreseeable future.

ence the sales item arrangement include the popularity distribution of different products, the more popular the more prominently and conveniently displayed, relationship between heights of average customers and the products that they most frequently purchase, right-handedness of most people, etc. These latter considerations distort the basic ontology, spawning its many views.

In our search for a fundamental ontology to be used in our Korean thesaurus, we have tried to construct an artefact of a typical Korean, someone whose world view most Koreans are expected to understand. Our goal is to be able to store away all Korean expressions in such a manner that a typical Korean speaker can guess/locate those expressions without too extensive a search. For our thesaurus project, we tested our still emerging ontology by trying to insert words/phrases from newspapers, magazines, and popular novels. We have adopted in constructing Lexaurus this same ontology that we have outlined from the above mentioned project. While such thesauri are great improvements over conventional dictionaries as tools for translators, organization of thesauri needs further improvement by isolating formal parameters and structuring word groups accordingly.

#### 4. Lexaurus and Ontology

Objects are organized according to some view of some category everywhere. Some typical examples are yellow pages of telephone books, various markets, scientific disciplines, object classes in object oriented programming language tools (e.g. Small Talk Browser), and Library of Congress classification of books. Similar classes of objects are grouped differently according to some particular needs at the time and place of the application in question. The ontology employed here for Lexaurus is a 'standard' ontology of an average Korean, obviously an artefact.

To demonstrate how ontology may vary from culture to culture and how different ontologies may affect the languages that represent the cultures, consider some relation terms. In-laws are in-laws in English but in a highly structured Korean society different kinds of in-laws need to be distinguished, for example, **brother-in-law**=two males married to sisters, 동서; husband to wife's brother depending on the age relation between wife and her brother, 매형 (if sister is older), 매제; husband to wife's sisters depend-

ing on the age relation between wife and her sister, 형부 (if wife is older), 제남; a male to his brother's wife depending on the age relation between the brothers 시숙 (if the husband is younger), 시동생.

The top level categories are: Human (body and senses, mind and ideas, feelings, behavior and will, language, society and institutions, values and ideals, occupations and crafts, sports and amusements, arts, science and technology), living things, natural phenomena, measure and shape, place and change of place, time and change of time. As we mention throughout this paper, an ontology is ephemeral to the extreme. One must fossilize an average person's ontology at its stablest moment and represent as sensibly and reasonably as one can.

We have started out by translating Roget's original ontology into Korean but then we quickly came to realize that Roget's organization is not maintainable. We thus have come to reorganize the top level as listed in the preceding paragraph. Next the second level of categories are assigned to one or more of these categories afresh. Finally lexical items are being grouped as colonies and then assigned to these second level categories. For instance, if one searches for an expression for a body of water of certain size, one looks for the place/change of place category, under which one finds body of water category. Under that category, one finds several colonies such as body of water that flows and body of water that is stationary. The body of water that flows category has words such as *stream, waterway, channel, river, brook, creek, rivulet, streamlet, rapid* and *tributary*.

Different organizations of these words within a colony may be viewed by the user according to their need. Parameters dictate the views' organizations. Macintosh operating system allows the user to view the files according to several parameters such as name, icon, date, and size of the files. In Lexaurus, each Korean reading under any English word entry can be viewed with one of the colonies it belongs to. Of course a word has many senses, which roughly correspond to our colonies. The entry often selects a particular sense out of possibly many senses the Korean reading is associated with. On those rare occasions when more than one sense still remains associated with the reading even under a given entry, the user will be allowed to choose the desired sense/colony. For that colony, the user now will take advantage of the view menu to select a parameter in accordance with which the colony in question will be displayed. Thus Lexaurus is just not

the list of entries with their basic meanings but rather a guide to the proper use of words as well.

## 5. Lexaurus and Dictionaries

As mentioned in the introductory section, bilingual dictionaries represent basically one-dimensional list of word entries, rarely expressing relationships among lexical items. For semantics an entry just lists a few sample target language expressions in a mostly arbitrary order. Lexaurus is a multidimensional and ontology-based bilingual dictionary. In a previous paragraph, we listed the parameters that we find useful in ordering words in colonies, which we repeat here for convenience, they are degrees of intensity, of formality, of technicality, of sincerity/jocularity, of familiarity, of generality, of vernacularity, of permanency, and of lexicalization, speaker's attitude toward the concept expressed, the speaker-hearer relation, discipline/specialty/area of use, and the gender of the speaker. Below we will illustrate some of these parameters that Lexaurus uses as a multidimensional dictionary.

**Degree of lexicalization.** Virtually all research papers/reports introduce certain concepts and their names and then use the names throughout the papers. If the research work gains a general acceptance, the names adopted in the paper might gain a broader acceptance beyond the readers of the work. People crystallize an expression when a concept becomes commonly popular and its description becomes necessary. Some expressions get lexicalized only among some specialists, never gaining general currency. Many motorists get to be irritated by rumble strips when they approach turnpikes so that they would slow down but very few notice the warning sign that contains the expression. This word 'rumble strip' has been translated as '진동구간' in one English-Korean dictionary and '전방의 위험을 운전자에게 차체의 진동으로 알리기 위해 도로 위에 잔 홈을 지어 포장한 부분' in another. The latter just gives an explanation but shies away from proposing any equivalent Korean word. The first offers a Korean word, unfamiliar to many Korean motorists, and yet understandable especially to those who recognize the underlying concept.

**Degree of intensity.** Consider as an example the colony of words having to do with wealthiness. How would one rank expressions such as

rolling in *money, well-wheeled, filthy rich, loaded, opulent, luxurious, flush, well-off, well-to-do, affluent, rich, wealthy, from hand to mouth, lean, poor, impoverished, beggarly, half-starved, starving, famished, wanting, poverty-stricken?* Or the colony of size words such as *big, huge, immense, vast, enormous, astronomic, jumbo, king-size, tremendous, prodigious, stupendous, macro, mega, giga, great big, Homeric, mighty, titanic, colossal, monumental, epic, towering, mountainous, monstrous, mammoth, mastodonic, gigantic, giant, gigantesque, Gargantuan, Herculean, elephantine, dinosaurian, whopping, walloping, whaling, whacking, spanking, full-sized, full-blown, oversize, sizable, grand, behemoth, leviathan, and bulky.* Degree of intensity suggests itself. To show the extent of surprise one hears people using expressions such as ‘*I almost jumped out of my skin*’ and ‘*my heart just stopped*’. How is it to be translated? The first has been translated in one novel as ‘어떻게 놀랐던지 녀이 있는대로 빠질 뻔했네’ We feel that an electronic dictionary must offer the service of colony extraction in some parameter-based ordering. A dictionary is written after all as a depository of information having to do with word use and meaning and one of the critical services expected from an electronic dictionary is to provide a range of expressions (colony) organized in some fashion so that the user may be guided through this organization to find the most appropriate, exact and desired expression.

**Formality.** Consider the following Korean expressions (colony) for ‘to die’: ‘*뺨다, 꺾다, 웨지다, 죽다, 가다, 염라대왕 만나다, 북망산으로 가다, 황천객이 되다, 운명하다, 세상을 떠나다(버리다), 돌아가시다, 사망하다, 서거하다 and 눈을 감다*’. English ‘croak’, for instance, would be closest to ‘뺨다’. Among all these words, only a few would be appropriate for funeral talks. A dictionary must assist the user to avoid any embarrassment by informing him of the degree of formality associated with each of these words.

**Speakers’ attitude.** Which of the following words would one employ to describe those that one favors and those that one disfavors: *corpulent, stout, fat, overweight, fattish, obese, adipose, gross, fleshy, beefy, meaty, hefty, porky, bloated, puffy, swollen, big-bellied, full-bellied, potbellied, gorbellied, plump, buxom, zaftig, full, huggy, rotund, fubby, stocky, stalwart, fat-assed* 뚱뚱한, 토실토실한, 미룩돼지 같은, 암사슴 같은,<sup>2</sup> 포동포동한, 비만한, 비대한, 미련하게(천치같이) 살이 쪼, 포실포실한.’ We feel that a bilingual dic-

<sup>2</sup> This seems to originate from the Bible, Song of Solomon.

tionary must match disfavored words in one language with disfavored ones in the other, never mixing favored ones with disfavored ones. One might admire, just like, dislike, hate, or even despise certain object. There is a whole range of attitudes one can have. A dictionary must have some means of informing the user of what attitude is associated with which word.

**Vernacularity.** From the colony having to do with ‘fatness’, consider ‘fubby (British)’ and ‘zaftig (yiddish).’ Koreans in the northern regions use expressions such as ‘상계도(아직도)’ or ‘전수(모두 합해서)’ that people from other regions do not. This presents a challenge in that the geographic or vernacular regions for one language differ by necessity from those for the other. How should the mapping go? In English, ‘Timbuktu’ is a word often used to denote remoteness and a similar expression is found virtually in all languages. The Korean word ‘무주구천동’ is perhaps similarly used for older Koreans who still remember that town to be a remote village when modern traffic facility was not yet available.

**Discipline/Specialty/Area of Use.** This is probably the most frequently used parameter in existing dictionaries. The entry for the English word ‘matrix’ has the following Korean equivalents in one of the dictionaries the authors looked at: (Anatomy) ‘모체’; (Biology) ‘세포간질’; (Mining) ‘맥석’; (Mathematics and computer) ‘행렬, 매트릭스’; (Printing) ‘자모, 모형, 발형’. We feel that a dictionary must inform the user as to what list of disciplines it uses.

**Other ontological categories.** Verbs and adjectives are often associated with certain semantic categories. For example, in one English-Korean dictionary, the entry for ‘mature’ lists the following: for living things, ‘완전히 성장한, 충분히 발육한’; for people, ‘성숙한, 분별력 있는’; for plans and ideas, ‘숙고한, 빈틈없는’; for government bonds, ‘만기의’. A full and mature ontology is needed so that such categorial matching may be done systematically. We are developing an ontology for the Korean language and the culture underlying it through our thesaurus project.

As with ‘rumble strip’, the semantics is fine one way but for those who are looking for ways of expressing the same in Korean the real task starts from this point onward. I do not know whether ‘울퉁불퉁(진동)구간’ is fairly widely accepted among Koreans, I rather doubt it, especially considering the fact that while the concept is well established in the US, I am not certain how firm the lexicalization has gotten even in this country. Exam-

ples are everywhere. 'Briefing' is well explained in any English-Korean dictionary but let us say we want to translate John Grisham's *'The Firm'* where a hoodlum is supposed to have discovered a situation critical in their little maneuver and is planning to 'brief' to his boss. No problem in concept but no English-Korean dictionary lists an actual Korean word to use that expresses the concept properly. The translator (KyungHee Kong) uses the Korean word '보고하다'. And we are saying that a lexically oriented dictionary should list this Korean word among others under the entry 'brief'.

## 6. Approaches to Data Collection for Lexaurus

Translators spend bulk of their translation time in identifying exact matches for words and phrases, expressions of varying sizes and units. And yet much of this work will be duplicated over and over again by others and often by the same translators themselves. Some translators develop skill in finding exact equivalents. Such skill will indeed distinguish those translators as experts. But this hard earned skill rarely benefits next generations of translators. There is a huge mine of resources to be tapped in this regard. For many decades, for many centuries in some cultures, translators have worked to find semantically and stylistically equivalent expressions between languages and left their heritage in their works. If we collect in a well organized dictionary such mappings from their scattered sources, translators will only need to select words they need from the list of the stored mappings and only in a rare occasion they may need to work to identify proper expressions on their own, not finding them in such a list. A collection of such mappings will serve many unique purposes, word bank for translators and resources for language teaching, to mention just a few.

Over many decades in Korean history, translations have been carried out and in the process collectively much time has been spent and efforts expended to find perfect matches. If we can collect the results of such efforts and make them available to users, duplicative time and efforts can be spared. Clearly much screening is needed to eliminate poor renderings and radical context dependencies, etc. One of the goals of this work is to bring translation out of art realm into business or at least science realm.

Field work is another approach that is needed. Trained workers are sent to different sites to collect target language expressions (e.g. hospitals and

doctors' offices for medical expressions, law firms for legal terms, computer labs for computerese, factories for everyday terms used there) and link them with corresponding English words. Eventually, the work will take a communal structure on a multinational scale. A distributed management over a network will become necessary due to the volatile nature and complexity of modern languages.

## 7. Tools for Implementing Lexaurus

**Merger.** We have been manually extracting such mappings by perusing magazine articles from *Newsweek*, *Readers' Digest*, *Watchtower*, and *Awake*, these being four major magazines available to the authors and translated into Korean, and many novels. After we create files of word mappings collected from such perusal, we have to screen off spurious mappings and duplications and eventually merge all collected mappings into one dictionary. The process of screening obviously cannot be done by a computer program but merging is. We have designed such a merger program, *Merger*, that takes a file entry at a time and compares it against the master file entries. If it finds the entry, then it peruses the listed items. If the same item is found, the program simply adds the source for that mapping to the source list attached to the item. If the item is not found under the entry, it then is added with appropriate marking on its index. If the entry is not found, it is created and the new mapping marked.

**Extractor and Clearing Experts.** We are working on automation of this mapping extraction process. The tools that are needed for our *Extractor* program include *Name Expert*, *Number Expert*, *Pronominal Expert*, *Sentential Expert* and *Morphology Expert*. *Name* and *Number* experts recognize proper names and numeric expressions and keep them out of the comparison field, reducing the list of words to be examined. Often *Pronominal Expert* does more than just eliminating a pronominal expression from the comparison arena. This is because translators frequently use for clarification purpose an expression they feel the pronoun stands for. *Sentential Expert* treats trashes or remnants, expressions that are there for various reasons but that are not worthy of listing as mapping in a dictionary. This includes prepositions, interjectives, connectives, particles, and some common verbs such as copula, 'have', and 'get'.

**Classifying Experts.** All these words that are excluded from consideration by various Experts are nevertheless needed for parsing purpose by Extractor. Noun Expert, Verb Expert, Adjective/Adverb Expert, Subject Expert, IC-Analyzer and Idiom Expert. These experts separate the remaining words into different categories to further assist the mapping. The morphology expert will do the needed morphology analyses and give the dictionary forms for both language expressions.

## 8. Example Extraction of Matches by Extractor

We will illustrate the utilities of these experts by two sentences from John Grisham's novel *The Firm* as is translated by KyongHee Kong (the first two sentences from chapter 5).<sup>3</sup>

*The small lobby outside Royce McKnight's office was empty when Mitch arrived precisely at eight-thirty, on schedule. He hummed and coughed and began to wait anxiously.*

‘미첼이 약속 시간에 맞춰 로이스 맥나이트의 사무실 밖의 자그마한 로비에 들어섰을 때 로비에에는 아무도 없었다. 미첼은 흠흠 소리를 내며 헛기침을 하고 짜증스럽게 기다렸다.

After the first group of experts complete their jobs, only a few words are left, marked below as bold-face letters.

*The **small lobby** outside Royce McKnight's office was empty when Mitch **arrived** precisely at eight-thirty, on schedule. He **hummed** and **coughed** and began to wait anxiously.*

미첼이 약속 시간에 맞춰 로이스 맥나이트의 사무실 밖의 자그마한 로비에 들어섰을 때 로비에에는 아무도 없었다. 미첼은 흠흠 소리를 내며 헛기침을 하고 짜증스럽게 기다렸다.

Eventually the second group of experts will first match groups of expressions as in (small lobby, 자그마한 로비), (on schedule, 시간에 맞춰), (Office, 사무실), (arrived, 들어섰을 때), (empty, 아무도 없었다),

<sup>3</sup> This illustration is somewhat idealized in that many problems needed for such performance are not resolved at this point of time.

(hummed, 흙흙소리를 내며), (coughed, 헛기침을 하고), (anxiously, 짜증스럽게), and (began to wait, 기다렸다). From these, Morphology Expert will extract the pairs: *anxiously* ‘짜증스럽게’; *arrive* ‘들어서다’; *cough* ‘헛기침을 하다’; *empty* ‘아무도 없다’; *hum* ‘흙흙거리다’; *lobby* ‘로비’; *office* ‘사무실’; *on schedule* ‘시간에 맞춰’; *small* ‘자그마한’; *wait* ‘기다리다’. Majority of pairings here (six out of ten to be exact) are not found in major full-size English-Korean dictionaries.

## 9. Utilities of Lexaurus

Such functionally adequate dictionary would be a revolutionary boost in language learning. Any alien resident who has lived in any foreign language/culture over some time realizes that conventional dictionaries are woefully inadequate for their everyday needs. A shopper in shopping malls, a patient discussing his symptoms with a doctor, a car driver talking with a mechanic concerning his broken automobile, and a parent wanting to have some serious discussion with a child, any individual in any practical language needs would quickly find that conventional dictionaries do not offer much service for them. Information they need they realize they must search somewhere else. Our lexaurus may prove to be a life skill depository that stores practical linguistic knowledge and that maintains such knowledge in a most accessible way.

Also if we discover all the parameters that go into selecting a particular set of expressions, when we discover all of them we can truly set out to automating, at least offering some genuine progress toward automating of the translation process. Instead of starting with blind proliferation of mappings of words, idioms, and phrases of sentences, the translator program can constrain such mappings first by selecting the context of use together with the intended denotations of the expressions and secondly by selecting their equivalents thus warding off much of exponential explosion encountered in blind uninformed attempts at automatic translation.

## References

- Goodsell, Don (1989) *Dictionary of Automotive Engineering*, London: Butterworths.

- Lindberg, D. A. B., B. L. Humphreys and A. T. McCray (1993) 'The Unified Medical Language System,' *Meth Inform Med.*, Vol 32. pp. 281-91.
- Meyer, Ingrid, Ynne Bowker and Karen Eck (1992) 'Cogniterm: an experiment in building a terminological knowledge base,' *Euralex 92*.
- Rector, A. L., W. D. Solomon, W. A. Nowlan and T. W. Rush (1995) 'A Terminology set for medical language and medical information systems,' *Meth Inform Med*.
- Rector, A. L., W. A. Nowlan and A. Glowinsky (1993) 'Goals for Conceptual Representation in the GALEN project,' In: C. Safran (ed). *SCAMC 93*, New York, McGraw-Hill 1993: pp. 414-418.
- Roget, Peter Mark (1992) *The Original Roget's International Thesaurus*, fifth edition. Robert L. Chapman, ed. Harper Perennial.
- Schmitz, Klaus-Dirk (1993) (ed) TKE 93: 'Terminology and Knowledge Engineering,' *Proceedings of the Third International Congress on Terminology and Knowledge Engineering*, Frankfurt/M.: Indeks Verlag.
- Skuce, Douglas (1993) 'A system for managing knowledge and terminology technical documentation,' In Klaus-Dirk Schmitz (ed) TKE 93: *Terminology and Knowledge Engineering: Proceedings of the Third International Congress on Terminology and Knowledge Engineering*, pp. 428-441.
- Skuce, Douglas (1995) 'A multi-functional knowledge management system,' *Knowledge Acquisition* (in print).
- Sowa, J. F. (1988) 'Using a lexicon of canonical graphs in a semantic interpreter,' in *Relational Models of the Lexicon-Representing Knowledge in Semantic Networks*, Sowa, J. F., ed. San Mateo (CA): Morgan Kaufmann.
- Sowa, J. F. (1991) 'Principles of Semantic Networks,' Sowa, J. F. ed. San Mateo (CA): Morgan Kaufmann.
- UMLS Knowledge Sources (1993) *Unified Medical Language System*, National Library of Medicine, US Department of Health and Human Services.
- UMLS (1991) *Fact Sheet*, Bethesda (Maryland, USA): Office of Public Information, National Library of Medicine.

## ABSTRACT

# Lexaurus: A Multidimensional, Ontology-based Bilingual Electronic Dictionary

Jonathan C. Oh and Hyo-Pil Shin

With the emerging technology through the computer resources, some of constraints we have accepted and lived with in connection with dictionaries yield places to some innovative projects. One such project is an electronic dictionary, which has words listed alphabetically but not limited to that ordering for information extraction. Lexaurus, a new type of dictionary, allows for extraction of groups of words according to some ontological categorizing, serving the roles of traditional dictionaries and thesauri as well. This paper outlines the major characteristics of our protocol system English-Korean lexaurus.

Jonathan C. Oh  
Computer Science  
University of Missouri at Kansas City  
Kansas City, MO 64110  
Fax 816-235-5159  
E-mail Oh@cstp.umkc.edu

Hyo-Pil Shin  
E-mail hpshin@cstp.umkc.edu