MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007


# PICO Linguist and BabelMeSH: Development and Partial Evaluation of Evidence-based Multilanguage Search Tools for MEDLINE/PubMed

**Paul Fontelo[1], Fang Liu[1], Sergio Leon[1], Abrahamane Anne[2], Michael Ackerman[1]**

*[1]High Performance Computing and Communications, Lister Hill National Center for Biomedical Communications*
*[2]National Library of Medicine Associate Fellow*
*National Library of Medicine, Bethesda, MD 20894*

**Abstract**

*PICO Linguist and BabelMeSH are multilanguage search tools intended for users whose native language is not English. A database of medical terms was created using concept identification equivalents of English terms to other languages. The primary sources of vocabularies were UMLS, MeSH, WHO EMRO and UMLF. The search interface changes according to the language selected which allows search terms to be entered in the native language. The user can limit the search output according to the language of publication but citations retrieved are in English only. Links may be provided to journals if published online. Evaluation of the French and Spanish versions using journal key words and a list of common diseases showed 77.5% and 86.5% accuracy respectively. User feedback was positive. PICO Linguist and BabelMeSH could be useful and convenient tools in finding current evidence sources in the medical literature especially for non-English medical terms that may be difficult to express in English.*

***Keywords:***

evidence-based medicine, multilanguage search, MeSH, MEDLINE/PubMed, PICO, UMLS

## Introduction

*ask*MEDLINE was developed as a free-text search tool for MEDLINE/PubMed.[1] We noticed that about 8% of queries early in its development were non-English terms. This is not unexpected because when searching the medical literature, it is often easier to use the language in which one is most accustomed to.[2] This observation motivated us to design search tools for MEDLINE/PubMed. PICO Linguist and BabelMeSH are the results of these efforts. Using their own native language could help find certain concepts that are difficult to translate into English. MeSH translations could be useful to healthcare providers and researchers who are not too comfortable with English in locating journal articles in MEDLINE/PubMed.

The PICO (Patient, Intervention, Comparison, Outcome) format is a method of searching medical literature that promotes the use of a focused, structured question. This search strategy leads to more precise searches[3] and facilitates the practice of evidence-based medicine using the well-built clinical question.[4] There are reports[5,6] in the literature of cross-language or multi-language projects for medical information retrieval but not many active Web sites. We propose PICO Linguist and BabelMeSH, concept-based search tools, for healthcare personnel and researchers who are not too familiar with English.

## Materials and methods

BabelMeSH is designed as a transparent multilanguage and cross-language interface. Users can submit medical terms in their native language (currently, Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian and Spanish) then a parser translates the query into English using a multi-language MySQL database. BabelMeSH then sends the query to PubMed through E-Utilities and returns English citations to the user.

PICO Linguist includes all the features that BabelMeSH have but instead of a single input box, PICO Linguist provides users with the structured PICO format consisting of Patient/Problem (P), Intervention (I), Comparison (C) and Outcome (O) input forms. The interface changes according to the input language selected.

### Multi-language database

The major source for most translation records in the databases is the UMLS Metathesaurus, which contains MeSH translations in French, German, Italian, Japanese, Portuguese, Russian and Spanish. Permissions were obtained from the contributing organizations. In addition, Dr. Stephan Darmoni and Dr. Patrick Ruch provided French MeSH translations and the unified medical lexicon for French (UMLF). Dr. Najeeb Al-Shorbaji, Regional Office for the Eastern Mediterranean, World Health Organization, kindly provided Arabic translations of MeSH and the Unified Medical Dictionary. Chinese terms were collected from multiple open source web sites.

The concept unique identifier (CUI) and its concept source were used to find translations from UMLS. Briefly, all the concepts in one non-English language and corresponding CUIs in UMLS were identified. If the English concept in UMLS from MeSH links to the same CUI, the English concept and the foreign language concept were paired.

Otherwise, the English concept from another vocabulary was selected.

A hierarchal scheme was devised in order to create separate translation tables for each foreign language using MySQL. The priority of source vocabulary from high to low in our system is: MeSH, UMLS-Meta, SNOMED and any other vocabulary. Table 1 illustrates the structure of each table, which includes CUI, English term, non-English term, and accented non-English term (if available).

*Table 1 – Translation table structure (French)*

| CUI | English | French | Accented French |
|---|---|---|---|
| C0702166 | Acne | Acne | Acné |
| C0019686 | HIV Antigens | Antigenes VIH | Antig_nes VIH |

For translation pairs that are not from UMLS where the foreign term and English term do not map to the same CUI, an internally generated concept ID is created by the system. An example is shown in Table 2 (P-45-52.)

### Character-based language processing

Character-based languages require special handling for string processing in Web applications and in the database. First, we unified character settings by encoding in UTF-8 (Unicode). The default character set in MySQL configuration file is UTF-8. Next, we installed the Multi-byte String Extension package for PHP because each letter or character in these languages may be represented by more than one byte in storage. Multi-byte String Extension can perform string processing for all of these languages. The character setting for internal encoding, HTML input and output in this extension is also in UTF-8.

### Parsing algorithm

The parser identifies terms found in the translation database and deletes all others not in the database. The corresponding English translations are sent to PubMed through E-utilities. For terms entered in the search form with accents, the parser searches the accented terms column in the database (Table 1.) If no matches are found, it will transform the accented input to the unaccented form, then search the database again and return the unaccented term. If no exact matches for some terms are found, BabelMeSH will suggest records that contain part of the non-English query. The user may modify the suggestions before sending to PubMed.

An algorithm was developed for complex searches such as combined MeSH terms or concepts. In order to compel the parser to find the optimal translation, the algorithm browses the input by two pointers, recursively (Figure 1.) First, translations of PubMed "stop words" are ignored. A multi-word input is then split into an array where each array element stores an input word. Two pointers (Pointer 1 and Pointer 2) are inserted at the beginning and the end of the array. Comparing them with the database identifies individual elements between the two pointers in the array. If no match is found, Pointer 2 moves towards the front of the array one element at a time, until a match is found, or Pointer 1 meets Pointer 2. Once a translation match for the multi-word term or phrase is found, these elements are removed from the array and Pointer 1 will move towards the end to the next unmatched element in the array. If two pointers meet before the translation processing finishes, Pointer 2 will reposition to the end of the array, and Pointer 1 will again move to the end, one element at a time. The translation search will repeat until Pointer 1 and 2 meets at the end of the array. Table 2 shows the records in the database related to the search illustrated in Figure 1.

*Table 2 – Records involved for translating the example query*

| CUI | English | Spanish |
|---|---|---|
| C0004916 | bed | cama |
| C0231290 | after | despues |
| C0442022 | lumbar | lumbar |
| C0033119 | puncture wound | puncion |
| C0553794 | spinal tap | puncion lumbar |
| C0004910 | bed rest | reposo en cama |
| P-42-52 | rest | reposo |

### User interface

An auto-complete feature (Figure 2, in Chinese) using XMLHttp object request JavaScript, provides suggestions as the query is entered. Suggestions can be selected from the drop down list.

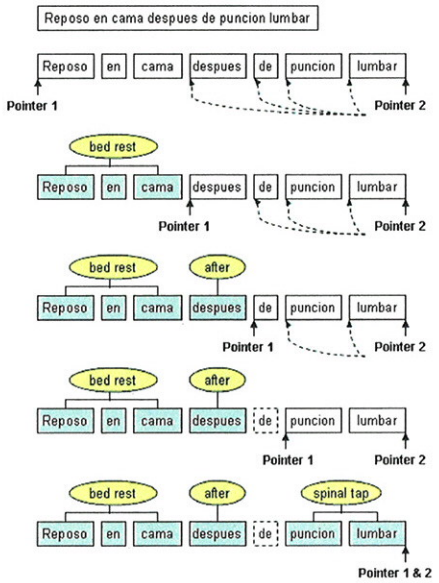


*Figure 1 – How the parser works*

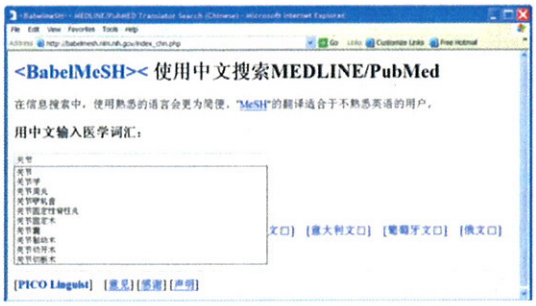


*Figure 2 – BabelMeSH (Chinese) with autosuggestion*

In PICO Linguist, the embedded JavaScript code makes the Web interface change dynamically with the input language choices, shown in Figure 3. The JavaScript code changes the text orientation to right to left when Arabic is selected (Figure 4.)


*Figure 3 – PICO Linguist English search page*

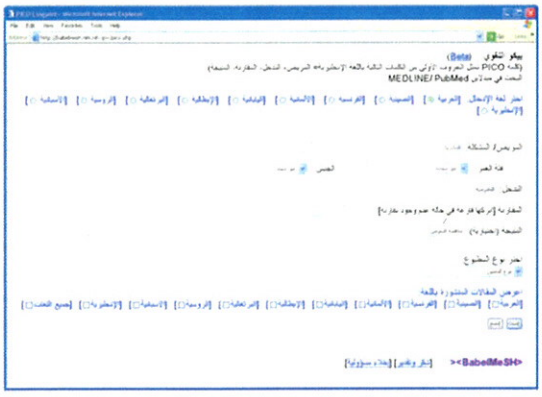


*Figure 4 – PICO Linguist interface in Arabic*

### Citation result filter

The user can select to view the titles of citations by language of publication. An output filter can display the results in one or more publication languages. The default is for articles published in all languages in PubMed journals. If a filter of publication language other than English is selected, the title of citation results will be shown in both English and the original language. However, this feature will only work for Latin languages. PubMed does not record original titles in character-based language, such as Arabic, Chinese, Japanese, and Russian.

BabelMeSH and PICO Linguist transmit the translated query to PubMed through E-Utilities. Retrieved citations are shown in English only because citations of foreign language journals indexed in MEDLINE are in English only. Links are provided to the foreign journal if published online. Journals may require subscription for full-text access. Search terms in the result's page are highlighted for the user's easy identification. Figure 5 demonstrates the result page of a query in Arabic.

### Evaluation

The full list of *BMJ Clinical Evidence* systematic reviews[7] was translated into Spanish by one of the authors (SL). These terms were searched in BabelMeSH then compared to the Spanish translations. The accuracy of BabelMeSH's English translation was compared to the original list. User opinion was obtained through a Web form. The evaluation of the French version consisted of three parts: (1) comparison of the author's own translation of keywords to English in journal articles published in French and the translations of the same terms to English by BabelMeSH; (2) actual user search terms from the our Web server's log files; (3) online user feedback questionnaire using a 5-point Likert scale to evaluate user opinion on the usefulness of BabelMeSH.

## Results

The search strategy in Figure 4 retrieved results shown below in Figure 5.

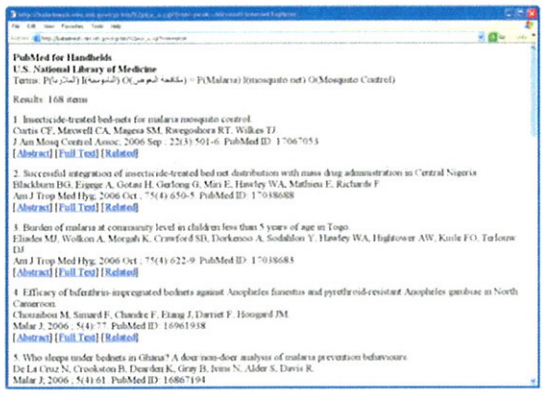


*Figure 5 – Retrieval of a PICO Linguist search in Arabic*

Table 3 shows the evaluation results of the French and Spanish versions of BabelMeSH and PICO Linguist. The two test sets used for evaluating the French version journal keywords and user query terms from server logs showed an accuracy of 75% and 79.9% respectively (mean=77.5%). Partial matches were obtained in 13.2% and 7.3% respectively. Full term matches and partial term matches equal to 88.2% and 87.2% of the test sets respectively. There were no matches in approximately 12% for both sets.

*Table 3 – Evaluation of the French and Spanish versions*

| | French | | Spanish |
|---|---|---|---|
| | Journal keywords | User query from server log | |
| **Total number of terms** | **174** | **179** | **221** |
| Exact or suggestions match | 130 | 143 | 191 |
| **% Total** | **75%** | **79.9%** | **86.4%** |
| Partial match | 23 | 13 | 20 |
| **% Total** | **13.2%** | **7.3%** | **9.0%** |
| Incorrect translation | 21 | 23 | 10 |
| **% Total** | **12%** | **12.8** | **4.5%** |

One hundred ninety one of the 221 terms from the disease list had exact matches for the Spanish version or 86.4% accuracy. For 20 terms (9%) partial matching was found but some translations missed one or two key words that could negatively affect a search. No translations were found for 10 (4.5%) search terms. We analyzed the complexity of search term or phrase as a measure of the parsing algorithm. The results (Table 4) show that accuracy is inversely related to the number of terms used.

Of the six responses received from the French feedback form, the average ratings (5=agree, 1=disagree) for the following statements were: 1) that BabelMeSH was useful 4.3/5; 2) the overall quality of citations retrieved was excellent 4.3/5; and 3) that they would continue to use BabelMeSH 4.6/5. All stated that they had previously searched MEDLINE in English and all except one declared that they would recommend it to others.

*Table 4 – Effect of search term count on translation accuracy*

| Number of terms | Total | Translated | % Total |
|---|---|---|---|
| 1 | 104 | 85 | 81.7 |
| 2 | 150 | 69 | 46 |
| 3 | 67 | 11 | 16.4 |
| 4 | 24 | 2 | 8.3 |
| 5 | 6 | 1 | 1.7 |
| 6 | 2 | 0 | 0 |

Comments from the Spanish users were mostly positive. Figure 6 shows the results of the two questions asked: (1) is searching MEDLINE/PubMed in your native language useful? and (2) how would you rate the results obtained? The mean responses were 4.3 and 3.9 respectively in a 5-point Likert scale.

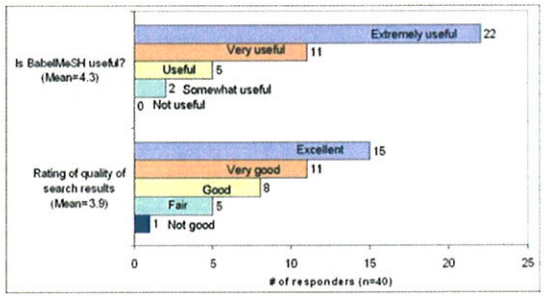


*Figure 6 - User feedback on Spanish version*

## Discussion

The aim of this project, as in most medical translations is to bring out the same cognitive "equivalent effect"[8] that an individual might be searching for in MEDLINE/PubMed. This is challenging in one language, even more when dealing with many languages as we have attempted here. The quality of sources is also a major determinant of success, one that we have no control over.

The evaluation of PICO Linguist and BabelMeSH demonstrates the usual challenges of medical translation due to the highly technical and scientific nature of medical language. The parsing algorithm is an attempt to overcome Newmark's "transparent collocation" phenomenon[8] as well as grammar. This is quite a challenge as shown in Table 4 where translation success drastically plummets as the complexity of the search term increases.

During development, we also found that the simple change in use of singular or plural nouns adversely affected the result of a query. This was common among all language versions. This is related to the MeSH translation in other languages, some translations only include descriptors but not all the valid entries. For example, the MeSH heading, "heart diseases", has 8 valid records, such as "heart disease" and "cardiac disease". For this medical concept, our database has differences among languages: six records in French, five in German, two in Japanese and one record each in Italian and Russian. If "heart disease", instead of "heart diseases", in Italian or Russian is entered, there will be no translation result. We have done some machine-normalization work on plural and non-plural words in those foreign languages, but it could also bring errors. MeSH contains more than 22,000 descriptors, but it also has greater than 130,000 additional valid entries that can help find the appropriate MeSH heading.

Table 3 shows the results of the evaluation of the French and Spanish versions. Using two different test sets, the results in French are comparable. Contrasted with the Spanish version the accuracy of translation was higher in Spanish, although this is really not a one-on-one comparison since they use entirely different databases. This is also

likely related to the size of the database, the Spanish database contains 788,835 entries while the French only 190,330.

We were unable to compare translation accuracy with other medical translation studies. Cross-language reports such as the one done by Volk,[9] evaluated their algorithms using precision and recall from MEDLINE searches. In this project we opted to use keywords and disease list to maximize the number of terms tested. It would be difficult to use precision and recall studies. We chose user satisfaction questionnaires to evaluate real-world searches.

Usability approval was encouraging. Thirty-eight responders in Spanish said that BabelMeSH was 'useful' to 'extremely useful' while only two found it 'somewhat useful' and none responded that it was not useful. The high rating on the quality of search results directly supports the cognitive equivalent effect of translations. This would have been possible only with accurate translations. Almost all of the 40 responders submitted comments that were even more enthusiastic like, "I've been waiting for it" and "There's nothing like mother language." There were also comments that expressed a preference to be able to read the citations and abstracts in their native language. The French evaluators were equally enthusiastic about these resources.

The highly positive feedback obtained from French and Spanish users showed us that there is need for multilanguage resources and the medical community may be eager to adopt utilities like PICO Linguist and BabelMeSH in their daily activities. We suspect the same type of response from users in other languages not tested in this phase of the project. Medical translation in not a simple matter, nevertheless, it is very important to have tools available for immediate use at the point-of-care when the healthcare professional is looking for current evidence in the field.

### Future challenges and developments

This resource is only as good as its database. It is a product of international collaboration and we invite collaborative research to make it even better. We also invite evaluation studies especially languages that are not covered by this report. The databases will be updated yearly as new editions of UMLS are released. Non-English translations of the abstract would be desirable. We will explore possible solution to this goal.

## Conclusion

PICO Linguist (http://babelmesh.nlm.nih.gov) and BabelMeSH (http://babelmesh.nlm.nih.gov/pico.php) provide templates for multilanguage search tools including character-based languages. They can be alternative search resources for healthcare personnel searching for current evidence in the medical literature for whom English is not their primary language. We invite collaborative research especially in improving its multilanguage database and evaluating its potential.

### Acknowledgments

We gratefully acknowledge colleagues[10] who provided advice, content and translation assistance for this project especially Dr. Najeeb Al-Shorbaji, EMRO, WHO.

## References

[1] http://askmedline.nlm.nih.gov
[2] Nelson SJ, Schopen M, Savage AG, et al. The MeSH translation maintenance system: structure, interface design, and implementation. Medinfo. 2004;11(Pt 1):67-9.
[3] Booth A, O'Rourke A, Ford N. Structuring the preseach reference interview: a useful technique for handling clinical questions Bull Med Libr Assoc. 2000 July; 88(3): 239246.
[4] Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence based decisions. ACP J Club 1995;123(3):A12–3.
[5] Boyer C, Baujard V, Scherrer JR. HONselect: multilingual assistant search engine operated by a concept-based interface system to decentralized heterogeneous sources. Medinfo. 2001;10(Pt 1):309-13.
[6] Hersh WR, Donohoe LC. SAPPHIRE International: a tool for cross-language information retrieval. Proc AMIA Symp 1998:673-7
[7] A full list of *BMJ Clinical Evidence* systematic reviews. http://www.clinicalevidence.org/ceweb/topics/topics-static.jsp
[8] Newmark P. A layman's view of medical translation. Br Med J. 1979 Dec 1;2(6202):1405-7.
[9] Martin Volk, Bärbel Ripplinger, Spela Vintar, Paul Buitelaar, Diana Raileanu and Bogdan Sacaleanu: Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval. In: International Journal of Medical Informatics. 67(1-3). 2002. 97-112.
[10] http://babelmesh.nlm.nih.gov/ac.php

### Address for correspondence

Paul Fontelo
High Performance Computing and Communications
National Library of Medicine
Bethesda, MD 20894
fontelo@nlm.nih.gov