MetaTrans
Multilingual Meta-Translator

Jan Pomikálek

Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
xpomikal@fi.muni.cz

Abstract. This paper presents MetaTrans, a meta-search engine for online dictionaries. With this software, users are able to find translations in a number of online dictionaries simultaneously. The MetaTrans features a web interface which is easy to use. The modular design of the tool enables adding support for more online dictionaries with minimal effort. MetaTrans also utilizes information from text corpora, WordNets and a morphological analyzer.

1 Introduction

There are many freely available online dictionaries on the web which contain large vocabularies for many language pairs. While common terms are often well covered by each of these dictionaries, the coverage of specialized terms and phrases differs from dictionary to dictionary. An obvious step for a user who did not find a translation of an unknown term in their dictionary is to look into another one. This may mean that the term needs to be searched for multiple times in many different dictionaries until a satisfying translation is found. This can be a tedious and time-consuming process.

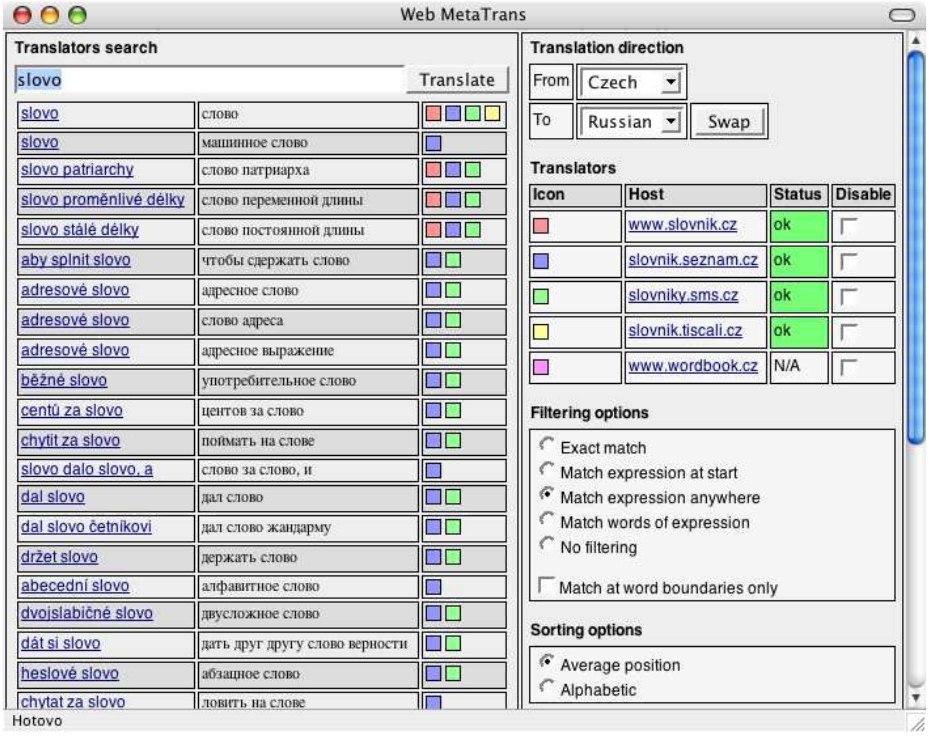
We present a system which acts as a meta-translator for a virtually unlimited number of online dictionaries. If a user looks for a translation of term X from language A into language B, the system operates as follows. First, a list of online dictionaries is determined which support translating from A to B. Then each of these dictionaries is searched for a translation of X. The results are merged and returned to the user.

Five online dictionaries are currently supported by the MetaTrans. However, as long as the design of the system is fully modular, it is very easy to add support for more online dictionaries.

The MetaTrans back-end is freely available on CPAN[1]. The MetaTrans web application can be accessed at http://metatrans.fi.muni.cz/ with some restrictions when used on computers outside Masaryk Universite. These restrictions are specified later in this paper.

[1] http://search.cpan.org/author/JANPOM/MetaTrans-1.04/bin/metatrans

Petr Sojka, Aleš Horák (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2007, pp. 109–115, 2007. © Masaryk University, Brno 2007



Translators search

slovo — Translate

slovo	слово	🟥🟦🟩🟨
slovo	машинное слово	🟦
slovo patriarchy	слово патриарха	🟥🟦🟩
slovo proměnlivé délky	слово переменной длины	🟥🟦🟩
slovo stálé délky	слово постоянной длины	🟥🟦🟩
aby splnit slovo	чтобы сдержать слово	🟦🟩
adresové slovo	адресное слово	🟦🟩
adresové slovo	слово адреса	🟦🟩
adresové slovo	адресное выражение	🟦🟩
běžné slovo	употребительное слово	🟦🟩
centů za slovo	центов за слово	🟦🟩
chytit za slovo	поймать на слове	🟦🟩
slovo dalo slovo, a	слово за слово, и	🟦
dal slovo	дал слово	🟦🟩
dal slovo četníkovi	дал слово жандарму	🟦🟩
držet slovo	держать слово	🟦🟩
abecední slovo	алфавитное слово	🟦
dvojslabičné slovo	двусложное слово	🟦🟩
dát si slovo	дать друг другу слово верности	🟦🟩
heslové slovo	абзацное слово	🟦🟩
chytat za slovo	ловить на слове	🟦

Hotovo

Translation direction

From: Czech

To: Russian — Swap

Translators

Icon	Host	Status	Disable
🟥	www.slovnik.cz	ok	☐
🟦	slovnik.seznam.cz	ok	☐
🟩	slovniky.sms.cz	ok	☐
🟨	slovnik.tiscali.cz	ok	☐
🟪	www.wordbook.cz	N/A	☐

Filtering options

○ Exact match
○ Match expression at start
● Match expression anywhere
○ Match words of expression
○ No filtering

☐ Match at word boundaries only

Sorting options

● Average position
○ Alphabetic

Fig. 1. MetaTrans main window

2 MetaTrans

MetaTrans is written in pure Perl using object oriented programming. For each online dictionary, a simple object class needs to be created which implements two methods – create_request and process_response.

The **create_request** method accepts three parameters – the term to be translated, the source language and the destination language. It returns a HTTP request which the given online dictionary will respond to with the appropriate list of translations. For example, to translate the word *pes* from Czech to English using the `slovnik.seznam.cz` online dictionary, the following HTTP GET request is required: `http://slovnik.seznam.cz/search.py?wd=pes&lg=cz_en`.

The **process_response** method also accepts three parameters – the response from the online dictionary, the source language and the destination language. The response is returned as an HTML page. The method is responsible for extracting the translations from the HTML code. This can be done either by using regular expressions or using an HTML parser. A relevant part of the response to the request specified above is shown in Fig. 2 The process_response method is responsible for converting it into the list of pairs which is shown in Fig. 3

Additional information is sometimes provided with a translation which specifies the senses of the terms. For the sake of consistency, this information should always be enclosed in parentheses. The process_response method is responsible for the appropriate conversion if necessary. In the previous example, the translations are already in the correct format. However, if we had something like *pes – lovecký* instead of *pes (lovecký)* we would need to convert it.

Apart from the two above mentioned methods, the object class also has to contain a constructor which sets the meta-data of the given online dictionary, such as name and list of supported language pairs.

```
<li> <strong> <a href="search.py?lg=cz_en&wd=pes">pes</a></strong> &nbsp;-&nbsp;
  <a href="search.py?lg=en_cz&wd=tyke">tyke</a> </li>
<li> <strong> <a href="search.py?lg=cz_en&wd=pes">pes</a></strong> &nbsp;-&nbsp;
  <a href="search.py?lg=en_cz&wd=pooch">pooch</a> </li>
<li> <strong> <a href="search.py?lg=cz_en&wd=pes%20%28loveck%C3%BD%29">pes
  (lovecký)</a></strong> &nbsp;-&nbsp; <a href="search.py?lg=en_cz&wd=hound">hound</a>
  <a href="./sound/A/A22051.WAV" title="Přehrát zvuk"><img src="http://1.im.cz/sl/repro.gif"
  width="13" height="13" alt="" class="repro" /></a> </li>
<li> <strong> <a href="search.py?lg=cz_en&wd=pes%20%28t%C3%A9%C5%BE%20p%C5%99en.%29">pes
  (též přen.)</a></strong> &nbsp;-&nbsp; <a href="search.py?lg=en_cz&wd=dog">dog</a>
  <a href="./sound/A/A1331.WAV" title="Přehrát zvuk"><img src="http://1.im.cz/sl/repro.gif"
  width="13" height="13" alt="" class="repro" /></a> </li>
```

Fig. 2. Response sample for `slovnik.seznam.cz`

With the object classes for the online dictionaries, it is already straightforward to search for translations. For a given language pair (A, B) and the input term X, each dictionary is queried for the list of translations of X. In order to speed up the process a new thread is spawned for each dictionary so that all of them can be queried at the same time. When all the responses are retrieved, the results are merged. If the same translation is obtained from multiple sources, it is only displayed once and it is associated with the list of dictionary icons, in which it was found. This is useful information for the user. The more sources of the translation the more likely it is that the translation is correct. If on the other hand, the translation is only found in a single dictionary, it may be that the translation is incorrect or inappropriate.

```
pes - tyke
pes - pooch
pes (lovecký) - hound
pes (též přen.) - dog
```

Fig. 3. The result of processing the response sample

2.1 Sorting

MetaTrans supports two types of sorting the retrieved translations – alphabetically and by average position in the source dictionaries. The alphabetical sorting uses the relevance of the left side to the searched term as the primary criterion. With the left side we refer to the term in the source language. The relevance is determined using the following three rules:

1. The terms which are identical with the searched term are the most relevant.
2. The terms which contain the searched term as a substring are more relevant than the ones which do not contain it (*doggie* is more relevant to *dog* than *hound* is).
3. The more words in the term the lower relevance (*big dog* is more relevant to *dog* than *big bad dog* is).

The terms within the groups with the same relevance are sorted alphabetically, primarily by the left side, secondarily by the right side.

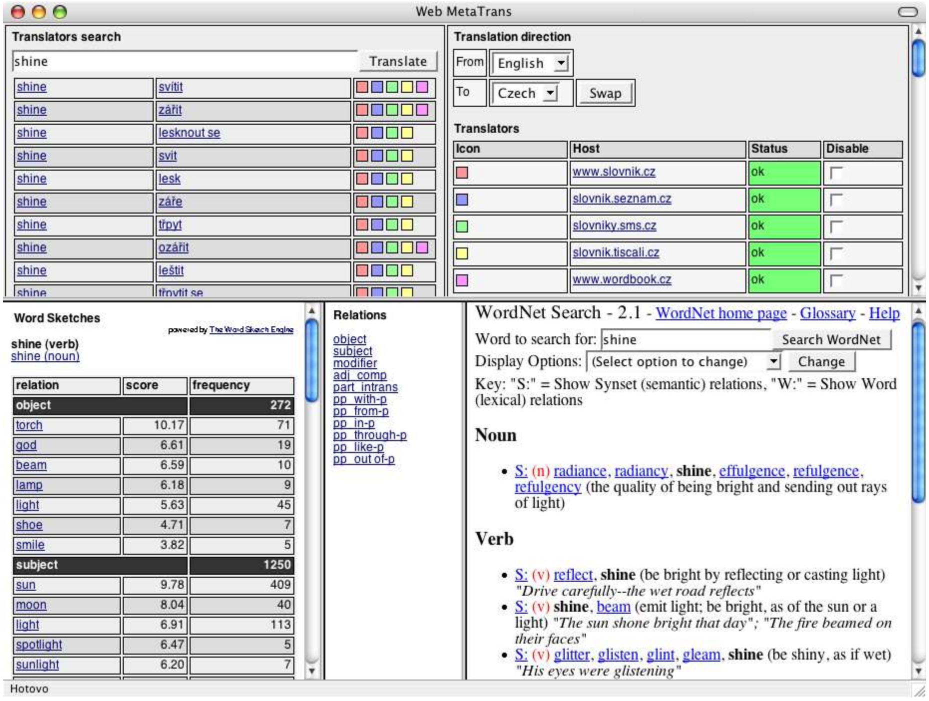
The next supported sorting is by average position and this is the default one. This is based on the assumption that the source dictionaries return the most usual or the most appropriate translations first (at the top). This is the most practical ordering for most users. The sorting by average position attempts to maintain this ordering. For each of the retrieved translations T, and for each source dictionary D, the position (rank) of the T is found within the translations returned by D. If the translation is not found in the list, the position behind the last translation returned by D is used. The positions are then averaged across the dictionaries used. The translations are sorted by the average position.

2.2 Filtering

Several filtering options are available in MetaTrans. The names of the options are mostly self-explanatory. We will therefore only demonstrate the effect of the most important ones on one example. Let us suppose that for a query *dog*, the translation of the following terms were found: *dog*, *doggie*, *dog bite*, *bad dog*, *bad doggie*, *hound*. The results of the filtering are presented in Table 1.

Table 1. The effects of filtering options for the input term *dog*

option	match at word boundaries	
	yes	no
exact match	dog	dog
match expression at start	dog bite	dog bite, doggie
match expression anywhere	dog, dog bite, bad dog	dog, doggie, dog bite, bad dog, bad doggie



Translators search

shine

shine	svítit	🟥🟦🟩🟨🟪
shine	zářit	🟥🟦🟩🟨🟪
shine	lesknout se	🟥🟦🟩🟨
shine	svit	🟥🟦🟩🟨
shine	lesk	🟥🟦🟩🟨
shine	záře	🟥🟦🟩🟨
shine	třpyt	🟥🟦🟩🟨
shine	ozářit	🟥🟦🟩🟨🟪
shine	leštit	🟥🟦🟩🟨
shine	třpytit se	🟥🟦🟩🟨

Translation direction

From: English

To: Czech

Translators

Icon	Host	Status	Disable
🟥	www.slovnik.cz	ok	□
🟦	slovnik.seznam.cz	ok	□
🟩	slovniky.sms.cz	ok	□
🟨	slovnik.tiscali.cz	ok	□
🟪	www.wordbook.cz	ok	□

Word Sketches

powered by The Word Sketch Engine

shine (verb)
shine (noun)

relation	score	frequency
object		272
torch	10.17	71
god	6.61	19
beam	6.59	10
lamp	6.18	9
light	5.63	45
shoe	4.71	7
smile	3.82	5
subject		1250
sun	9.78	409
moon	8.04	40
light	6.91	113
spotlight	6.47	5
sunlight	6.20	7

Relations

object
subject
modifier
adj_comp
part_intrans
pp_with-p
pp_from-p
pp_in-p
pp_through-p
pp_like-p
pp_out of-p

WordNet Search - 2.1 - WordNet home page - Glossary - Help

Word to search for: shine

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- S: (n) radiance, radiancy, **shine**, effulgence, refulgence, refulgency (the quality of being bright and sending out rays of light)

Verb

- S: (v) reflect, **shine** (be bright by reflecting or casting light) *"Drive carefully--the wet road reflects"*
- S: (v) **shine**, beam (emit light; be bright, as of the sun or a light) *"The sun shone bright that day"; "The fire beamed on their faces"*
- S: (v) glitter, glisten, glint, gleam, **shine** (be shiny, as if wet) *"His eyes were glistening"*

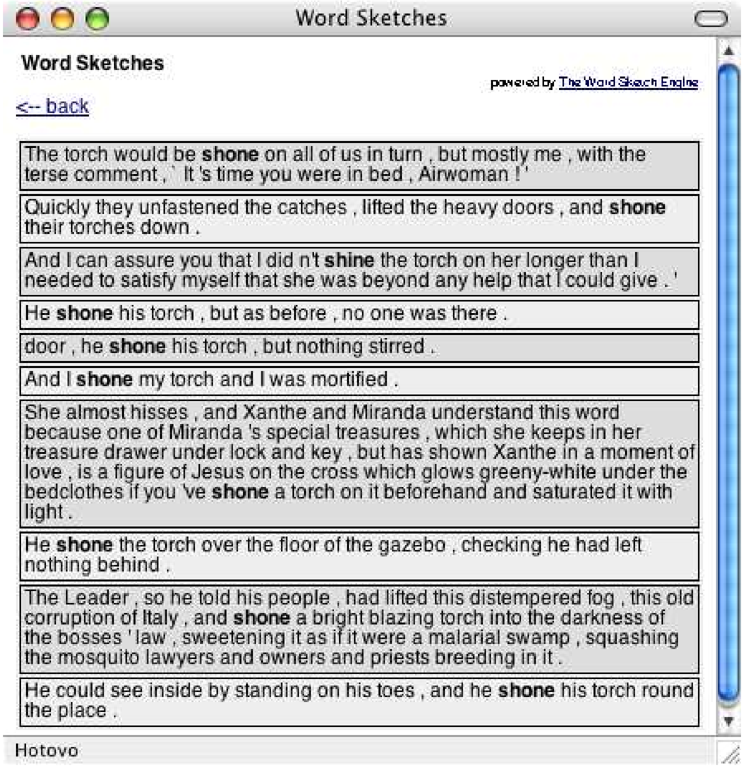
Fig. 4. Word sketches and WordNet information for *shine*

2.3 Additional Resources

MetaTrans can display additional information of various kinds for the retrieved translations. If the user clicks on a term, they are presented with the Word Sketches for the word, with the information from WordNet and with the morphological analysis of the word. Each of these resources is, however, only available for a limited number of languages.

A word sketch [1] is a summary of a word's grammatical and collocational behavior. Word sketches can be produced automatically from large annotated corpora using the Sketch Engine [2]. The collocates in the word sketches table displayed by the MetaTrans can be clicked on in order to display concordances from a text corpus (see Fig. 5). The concordances serve as usage examples of the given word in relation with the collocate.

The word sketches in the MetaTrans are available for Czech, English and French. The British National Corpus [3] is used for English, the Czech National Corpus for Czech. A web derived corpus is used for French. Since the last release of MetaTrans, word sketch tables were developed for a number of additional languages. These will be utilized in the next MetaTrans version. Due to license restrictions, the word sketches information is only available if the MetaTrans is accessed from the Masaryk University network.



Word Sketches

powered by The Word Sketch Engine

The torch would be **shone** on all of us in turn , but mostly me , with the terse comment , ` It 's time you were in bed , Airwoman ! '

Quickly they unfastened the catches , lifted the heavy doors , and **shone** their torches down .

And I can assure you that I did n't **shine** the torch on her longer than I needed to satisfy myself that she was beyond any help that I could give . '

He **shone** his torch , but as before , no one was there .

door , he **shone** his torch , but nothing stirred .

And I **shone** my torch and I was mortified .

She almost hisses , and Xanthe and Miranda understand this word because one of Miranda 's special treasures , which she keeps in her treasure drawer under lock and key , but has shown Xanthe in a moment of love , is a figure of Jesus on the cross which glows greeny-white under the bedclothes if you 've **shone** a torch on it beforehand and saturated it with light .

He **shone** the torch over the floor of the gazebo , checking he had left nothing behind .

The Leader , so he told his people , had lifted this distempered fog , this old corruption of Italy , and **shone** a bright blazing torch into the darkness of the bosses ' law , sweetening it as if it were a malarial swamp , squashing the mosquito lawyers and owners and priests breeding in it .

He could see inside by standing on his toes , and he **shone** his torch round the place .

Fig. 5. Contexts for *shine* in relation with *torch* from the BNC

WordNet [4] is a well-known lexical database which was originally developed for English. In the EuroWordNet project, WordNets were produced for several European languages and linked together. MetaTrans provides web access to the English, Czech and French WordNets.

For Czech words, morphological analysis is available which is provided by the web interface of the morphological analyzer *ajka* [5].

3 Conclusion

We have presented the multilingual meta-translator MetaTrans, a system for parallel searching in multiple online dictionaries. It has a convenient web interface which is easy to use especially for non-technical people. Its parallel processing of the online resources makes MetaTrans reasonably fast. Modular design enables using additional online dictionaries with very little effort. MetaTrans is completely language independent and can be used for translating between any pair of languages, as long as an online dictionary exists for the language pair.

Apart from the data from the online dictionaries, different kinds of language resources are used, such as word sketches, WordNets and a morphological analyzer. This makes MetaTrans unique resource for a wide range of users. Its large vocabulary also makes MetaTrans a valuable dictionary for obtaining specialized terminology.

Acknowledgments

This work has been partly supported by the Ministry of Education of CR within the Center of basic research LC536 and in the National Research Programme II project 2C06009.

References

1. Kilgarriff, A., Tugwell, D.: Sketching words. Lexicography and natural language processing: a festschrift in honour of B. TS Atkins, Euralex (2002) 125–137.
2. Kilgarriff, A., Rychlý, P., Smrž, P., Tugwell, D.: The Sketch Engine. Proceedings of Euralex (2004) 105–116.
3. Aston, G., Burnard, L.: The BNC handbook: Exploring the British National Corpus with SARA. Edinburgh University Press (1998).
4. Miller, G.: WordNet: A Lexical Database for English. Communications of the ACM **38**(11) (1995) p. 39.
5. Sedláček, R., Smrž, P.: A New Czech Morphological Analyser ajka. Proceedings of the TSD (2001) 100–107.