Synset Based Multilingual Dictionary: Insights, Applications and Challenges

Rajat Kumar Mohanty[1], Pushpak Bhattacharyya[1],
Shraddha Kalele[1], Prabhakar Pandey[1],
Aditya Sharma[1], Mitesh Kopra[1]

[1] Department of Computer Science and Engineering
Indian Institute of Technology Bombay, Mumbai - 400076, India
{rkm, pb, shraddha, pande, adityas, miteshk}@cse.iitb.ac.in

Abstract. In this paper, we report our effort at the standardization, design and partial implementation of a multilingual dictionary in the context of three large scale projects, *viz.*, (i) Cross Lingual Information Retrieval, (ii) English to Indian Language Machine Translation, and (iii) Indian Language to Indian Language Machine Translation. These projects are large scale, because each project involves 8-10 partners spread across the length and breadth of India with great amount of language diversity. The dictionary is based not on words but on wordnet SYNSETS, *i.e.*, concepts. Identical dictionary architecture is used for all the three projects, where source to target language transfer is initiated by concept to concept mapping. The whole dictionary can be looked upon as an *M X N* matrix where *M* is the number of synsets (rows) and *N* is the number of languages (columns). This architecture maps the lexeme(s) of one language- standing for a concept- with the lexeme(s) of other languages standing for the same concept. In actual usage, a preliminary WSD identifies the correct row for a word and then a lexical choice procedure identifies the correct target word from the corresponding synset. Currently the multilingual dictionary is being developed for 11 languages: *English, Hindi, Bengali, Marathi, Punjabi, Urdu, Tamil, Kannada, Telugu, Malayalam* and *Oriya*. Our work with this framework makes us aware of many benefits of this multilingual concept based scheme over language pair-wise dictionaries. The pivot synsets, with which all other languages link, come from Hindi. Interesting insights emerge and challenges are faced in dealing with linguistic and cultural diversities. Economy of representation is achieved on many fronts and at many levels. We have been eminently assisted by our long standing experience in building the wordnets of two major languages of India, *viz.*, Hindi and Marathi which rank 5th (~500 million) and 14th (~70 million) respectively in the world in terms of the number of people speaking these languages.

Keywords: Multilingual Dictionary, Dictionary Standardization, Concept Based Dictionary, Light Weight WSD and Lexical Choice, Multilingual Dictionary Database

1 Introduction

In any natural language application, dictionary look-up plays a vital role. We report a model for multilingual dictionary in the context of large scale natural language processing applications in the areas of Cross Lingual IR and Machine Translation. Unlike any conventional monolingual or bilingual dictionary, this model adopts the *Concepts* expressed as wordnet synsets as the pivot to link languages in a very concise and effective way. The paper also addresses the most fundamental question in any lexicographer's mind, *viz., how to maintain lexical knowledge, especially in a multilingual setup, with the best possible levels of simplicity and economy*? The case study of multiple Indian languages with special attention to three languages belonging to two different language groups (such as, *Germanic* and *Indic*) within the Indo-European family - English, Hindi and Marathi- throws lights on various linguistic challenges in the process of dictionary development.

The roadmap of the paper is as follows. Section 2 motivates the work. Section 3 is on related work. The proposed synset based model for multilingual dictionary is presented in section 4. Section 5 is on how to tackle the problem of correct lexical choice on the target language side in an actual MT situation through a novel idea of word alignment. Linguistic challenges are discussed in Section 6. Creation, storage and maintenance of the multilingual dictionary is an involved task, and the computational framework for the same is described in section 7. Section 8 concludes the paper.

2 Motivation

Our mission is to develop a single multilingual dictionary for all Indic languages plus English in an effective way, economizing on time and effort. We first discuss the disadvantages of language pair wise conventional dictionaries.

2.1 Disadvantages of Conventional Bilingual Dictionaries

In a typical bilingual dictionary, a word of L_1 is taken to be a lexical entry and for each of its senses the corresponding words in L_2 are given. It is possible that one sense of W_i in L_1 is exactly the same as one of the senses of W_j in L_1. This means that W_i and W_j are synonymous for a given sense. An example of this is dark and evil where one of the senses of dark and evil overlaps as for example in dark deeds and evil deeds. This phenomenon is abundant in any natural language. In a conventional dictionary, there is no mechanism to relate W_i with W_j in L_1, though they conceptually express the same meaning. In turn, the corresponding words for W_i and W_j in L_2 are no way related to each other though conceptually they are. That is a major drawback, because of which conventional pair wise dictionaries cannot be used effectively in natural language application, especially when multiple languages are involved.

The other disadvantage of the conventional dictionary is the duplication of manual labor. If an MT system is to be developed involving *n* languages, *n(n-1)/2* language pair wise dictionaries have to be created. For instance, if we consider *6* languages, *30* bilingual dictionaries have to be constructed. Additionally will be required *15* perfect bilingual lexicographers- by no means an easy condition to meet.

Finally, the effort of incorporating semantic features in *O(n2)* dictionaries is duplicated by *n/2* lexicographers- a wastage of manual labor and time.

3 Related Work

Our model has been inspired by the need to efficiently and economically represent the lexical elements and their multilingual counterparts. The situation is analogous to Eurowordnet [1] and Balkanet [2] where synsets of multiple languages are linked among themselves and to the Princeton Wordnet ([3], [4]) through Inter-lingual Indices (ILI). Our framework is similar, except for a crucial difference in the form of cross word linkages among synsets (explained in section 5). Another difference is that there are semantic and morpho-syntactic attributes attached to the concepts and their word constituents to facilitate MT. The Verbmobil project [5] for speech to speech multilingual MT had pair wise linked lexicons. To the best of our knowledge, no major machine translation nor CLIR project involving multiple large languages has ever used concept based dictionaries.

The framework has indeed been motivated by our creation of the Marathi Wordnet [6] by transferring from the Hindi Wordnet [7]. We noticed the ease of linking the concepts when two languages with close kinship were involved ([8], [9]).

4 Proposed Model: Concept-based Multilingual Dictionary

We propose a model for developing a single dictionary for n languages, in which there are linked concepts expressed as synsets and not as words. For each concept, semantic features- which are universal- are worked out only once. As for morph-syntactic features, their incorporation will demand much less effort, if languages are grouped according to their families; in other words we can take advantage of the fact that close kinship languages share morpho-syntactic properties. Table 1 illustrates the concept-based dictionary model considering three languages from two different families.

Table 1. Proposed multilingual dictionary model

Concepts	***L*₁ (English)**	***L*₂ (Hindi)**	***L*₃ (Marathi)**
Concept ID: Concept description	(W_1, W_2, W_3, W_4)	(W_1, W_2, W_3, W_4, W_5 W_6, W_7, W_8)	(W_1, W_2, W_3, W_4, W_5 W_6, W_7, W_8, W_9, W_{10})
02038: a typical star that is the source of light and heat for the planets in the solar system	(sun)	(सूर्य, सूरज, भानु, दिवाकर, भास्कर, प्रभाकर, दिनकर, रवि, आदित्य, दिनेश, सविता, पुष्कर, मिहिर, अंशुमान, अंशुमाली)	(सूर्य, भानु, दिवाकर, भास्कर, प्रभाकर, दिनकर, मित्र, मिहिर, रवि, दिनेश, अर्क, सविता, गभस्ति, चंडांशु, दिनमणी)
04321: a youthful	(male_child,	(लड़का, बालक, बच्चा,	(मुलगा, पोरगा, पोर,

male person	boy)	छोकड़ा, छोरा, छोकरा, लौंडा)	पोरगे)
06234: a male human offspring	(son, boy)	(पुत्र, बेटा, लड़का, लाल, सुत, बच्चा, नंदन, पूत, तनय, तनुज, आत्मज, बालक, कुमार, चिरंजीव, चिरंजी)	(मुलगा, पुत्र, लेक, चिरंजीव, तनय)

Given a row, the first column is the pivot for n number of languages describing a concept. Each concept is assigned a unique ID. The columns (2-4) show the appropriate words expressing the concepts in respective languages. To express the concept ‘04321: a youthful male person’, there are two lexical elements in English, which constitute a synset. There are seven words in Hindi which form the Hindi synset, and four words in Marathi which constitute the Marathi synset for the same concept, as illustrated in Table 1. The members of a particular synset are arranged in the order of their frequency of usage for the concept in question. The proposed model thus defines an *M X N* matrix as the multilingual dictionary, where each row expresses a concept and each column is for a particular language.

4.1 Advantages of the concept-based multilingual dictionary

(a) The first advantage of the proposed model is economy of labor and storage. Semantic features like *[±Animate, ±Human, ±Masculine, etc.]*, are assigned to a nominal concept and not to any individual lexical item of any language. Similarly, the semantic features, such as *[+Stative (e.g., know), +Activity (e.g., stroll), +Accomplishment (e.g., say), +Semelfactive (e.g., knock), +Achievement (e.g., win)]* are assigned to a verbal concept. These semantic features are stored only once for each row and become applicable independent of any language. Consequently, lexical entries with highly enriched semantic features can be added to a dictionary for as many languages as required within a short span of time.

(b) The dictionary developed in this approach also serves all purposes that either a monolingual or bilingual dictionary serves. A monolingual or bilingual dictionary can automatically be generated from this concept-based multilingual dictionary. The quality of such monolingual or bilingual dictionaries is better than that of any conventional bilingual dictionary in terms of lexical features.

(c) The model admits of the possibility of extracting a domain specific dictionary for all or any specific language pair. This is because the synsets or concepts pertaining to a domain can be selected from among the rows in the *M X N* concepts *vs.* languages matrix.

(d) The language group which lacks competence in the pivot language- which in our case is Hindi- can benefit from the already worked out languages. It may be the case that the lexicographers of language L_6 do not have enough competence in the pivot language L_{pivot}. They can look for a language Ln which they are comfortable with and use L_n as pivot to link L_6. This paves way for the seamless integration of a new language into the multilanguage dictionary.

5 Word-Alignment in the Proposed Model

In an actual MT situation, for every word or phrase in the source language a single word or phrase in the target language will have to be produced. The multilingual dictionary proposed by us links concepts which are sets of synonymous words. This is a major difference from the conventional bilingual dictionary in which a word *(SW_1)* in the source language is typically mapped to one or more words in the target language depending upon the number of senses SW_1 has. This implies that for each sense of SW_1, there is a single target language word TW_1. In our concept-based approach, even if we choose the right sense of a word in the source language *(SW_1)*, there is still the hurdle of choosing the appropriate target language word. This lexical choice is a function of complex parameters like situational aptness and native speaker acceptability. For example, the concept of ‘the state of having no doubt of something’ is expressed through the Hindi synset having six members (निश्शंक, अनाशंकित, आशंकाहीन, बेखटक, बेफ़िक्र, संशयहीन) and through the Marathi synset having four members (निःशंक, निर्धास्त, निर्भ्रांत, शंकारहित). However, the third member in the Hindi synset आशंकाहीन is appropriately mapped to the fourth member in the Marathi synset शंकारहित. Though the mapping of the third member in the Hindi synset (i.e., आशंकाहीन) with the first member of the Marathi synset (*i.e.*, निःशंक) expresses the same meaning, this substitution sounds quite unnatural to the native speakers.

We tackle the problem of correct lexical choice on the target language side by proposing a novel approach of word-alignment across the synsets of languages. Word-alignment refers to the mapping of each member of a synset with the most appropriate member of the synset of another language. For instance, when the word लड़का ‘boy’ in Hindi in the sense of ‘a young male person’ needs to be lexically transferred to Marathi, there are four choices available in the synset, as illustrated in Figure 1.

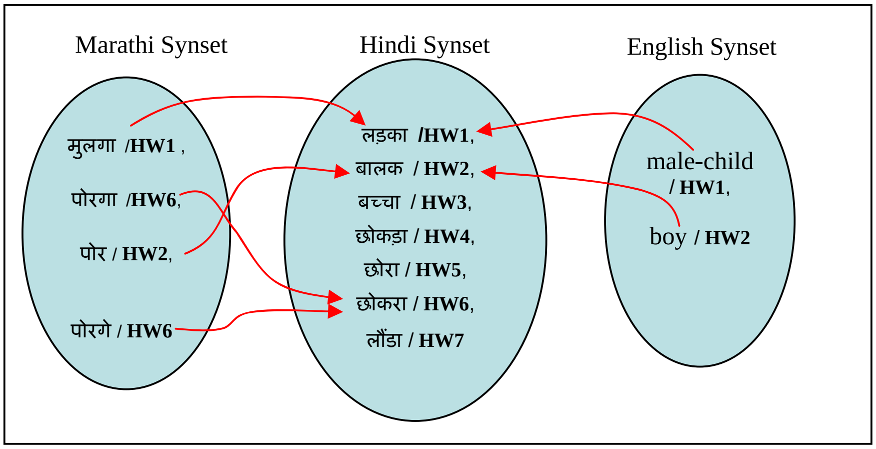


Fig. 1. Illustration of aligned synset members for the concept: *a youthful male person*

Considering Hindi as the pivot, we propose that each of the four words in the Marathi synset be linked to the appropriate Hindi word in the direction Marathi→Hindi and each of the two words in English synset has to be linked with the appropriate Hindi word in the direction English→Hindi. As a result, the first and the third member of the Marathi synset (*i.e.,* मुलगा and पोर) are mapped to two different Hindi words (*i.e.,* मुलगा→लड़का, पोर→बच्चा). The second and the fourth member in Marathi synset are linked to one word (*i.e.,* पोरगा→छोकरा and पोरगे→छोकरा) in the Hindi synset. Three words in Hindi synset (*i.e.,* HW4, HW5, HW7) are left without being linked, as shown in Figure 1. In a situation, when a Marathi word is aligned with a single Hindi word (*e.g.,* मुलगा → लड़का) for a particular concept in the direction of Marathi→Hindi, from our past experience we assume that the lexical transfer in the reverse direction (Hindi→ Marathi) also holds good, yielding लड़का → मुलगा.

Following this strategy of alignment of synset members of Marathi (or any other language) with the synset members of the pivot (*i.e.,* Hindi in the present scenario), we are having four types of situation to perform a lexical transfer from any language to any language:

Situation (1) One-to-One
Situation (2) Many-to-One
Situation (3) One-to-Many
Situation (4) No link

In situation (1), the source word is found to be linked to a single target word, via a synset member of the pivot if it is neither the source nor the target for any lexical transfer. For instance, the Marathi word मुलगा can be transferred to the Hindi target word लड़का, and the Marathi word मुलगा can be transferred to the English target word '*boy*' via the pivot Hindi word लड़का. In situation (1), virtually there is no problem in performing the lexical transfer maintaining the best naturalness to the target language speakers. In situation (2), two words from the source language synset are linked to a single word in target language, *e.g.,*पोरगा → छोकरा and पोरगे → छोकरा. Hence, there is no issue involved in lexical transfer maintaining the naturalness. The situation (3) arises when the pivot is taken as the source language in any practical application, *e.g., Hindi→ Marathi*. The lexical transfer involves a puzzle with respect to the naturalness of the target word. Since the members of a synset are ordered according to their frequency of usage for a concept, we are inclined to choose the first member of the target synset as the best in this situation. For instance, the source Hindi word छोकरा '*boy*' has two choices in the target Marathi synset, *i.e.,* पोरगा and पोरगे, as shown in figure 1. Since पोरगा appears prior to पोरगे, we choose पोरगा for lexical transfer. In situation (4), where *no link* is available between the source word and the target word, we choose the first member of the target synset for lexical transfer. If we need to transfer the Marathi word पोर to English, there is no consecutive link available, since it stops at बच्चा/HW3 in the pivot (*cf. figure 1*). However, we choose the first member of the English synset, *i.e., boy* for Marathi पोर, which is quite appropriate and widely acceptable. Similarly, if the English word *boy* happens to be the source in the sense of '*a youthful male person'*, the first member of the Marathi synset (*i.e.,* मुलगा) is chosen as the target for lexical transfer, even if its link stops at बालक/HW2 in the pivot (*cf.*

figure 1). In section 8, we present a user-friendly tool to align the members of the synsets across languages with respect to a particular concept. We also present a lexical transfer engine to make the aligned data usable in any system.

6 Linguistics Challenges Involved

In the process of synset based multilingual dictionary development, we face a number of challenges to deal with linguistic and cultural diversity. In this section, we present a few cases that we experienced while dealing with three languages, *i.e.*, English, Hindi and Marathi.

(a) A concept may be expressed using different syntactic category in different languages. For example, the nominal concept कलौंजी *'stuffed vegetable'* in Hindi is expressed through an adjectival concept भरली *'stuffed'* in the expression भरलेली भाजी *'stuffed vegetable'* in Marathi.

(b) It is often the case that a concept is expressed through a synthetic expression in one language, but through a single word expression in the other language. For example, the concept *'reduce to bankruptcy'* is expressed through a single word in English but through a synthetic expression in Hindi and Marathi, as illustrated in Table 2.

Table 2. Illustration of *single word vs. synthetic expressions*

Concept	English	Hindi	Marathi
'reduce to bankruptcy'	bankrupt *(V)*	दिवाला निकालना *(N+V)* *'to make bankrupt'*	दिवाळे काढणे *(N+V)* *'to make bankrupt'*
'resulting from careful thought'	considered *(ADJ)*	विचारपूर्वक किया हुआ *(ADV+VERB)* *'thoughtfully done'*	विचारपूर्वक केलेला *(ADV+VERB)* *'thoughtfully done'*
'least in age than the other person'	youngest *(ADJ)*	कनिष्ठ *(ADJ)*	सर्वात लहान *(N+ADJ)* *'among-all less-in-age '*

Considering Hindi as the pivot in the process of dictionary development in our approach, one has to deal with two kinds of situation: (i) synthetic expression in the pivot to a single word expression in the other language, (ii) single word expression in the pivot to a synthetic expression in the other language. In situation (ii), the question arises with respect to its morpho-syntactic category in the dictionary. Because, the synthetic element is often constituted of different syntactic categories, as shown in Table 2. In such a situation, we consider the grammatical function of the synthetic element and assign the category accordingly. For example, the Marathi expression विचारपूर्वक केलेला *(ADV+VERB) 'thoughtfully done'* refers to an adjectival function at the grammatical level, hence its syntactic category is assigned as '*adjective*'.

(c) When a word expressing the meaning specific to a particular language and culture has to be mapped to another language in the dictionary, we find two ways to express the concept in another language: *(i) using a synthetic expression*, *(ii) using transliteration*, if the synthetic expression is larger. For example, the culture specific concept of '*ornaments and other gifts given to the bride by the bridegroom on the day of wedding'* is lexicalized in Hindi yielding चढ़ावा, but a Marathi speaker has to use a larger synthetic expression विवाहसमयी वराकडून वधुला दिले जाणारे दागिने '*at-the-time-of-wedding–bridegroom–bride– given–ornament'* to express the same concept. The Hindi word सेहरा '*garland*' is a culture specific word which has no lexical equivalent in Marathi. Even using a large synthetic expression does not express the borrowed concept naturally. In such a situation, we transliterate the culture specific word into Marathi.

It is also the case that a concept is culture specific to a language other than the pivot. For example, the Marathi culture specific concept, *e.g.*, माहेरवाशीण '*a woman who has come to stay at her parents' place after her marriage'*, is not expected to be available in the pivot language dictionary in the initial phase. Therefore, such culture specific concepts are added to the Marathi dictionary in a monolingual manner without being mapped to the pivot language. But those are marked for review using the dictionary development tool. At a later phase, each language specific cultural concepts can be taken, and systematically added to the pivot language to enrich the pivot, and in turn, the whole multilingual dictionary with multicultural concepts.

(d) Given Hindi as the pivot language, when we develop and link the Marathi dictionary, we come across a strange situation. A concept initially recorded in Hindi dictionary, having a singleton member in the pivot synset, can be expressed through more than one finer concept in Marathi. The Hindi word फ़ीका means '*the food prepared with less sugar, salt or spice'*, the equivalent of which is expressed in Marathi through three distinct words expressing three distinct finer concepts, *i.e.*, अगोड '*less sweet'*, अळणी '*less salty*, and मिळमिळत '*less spicy'*. These three words cannot be taken as the members of a single synset in Marathi for the concept '*the food prepared with less sugar, salt or spice'*, since the three-way finer meaning distinction is very natural to Marathi speakers. Had it been the case that Marathi were the pivot, we could have been tempted to add three different concepts into Marathi dictionary, and in turn, Hindi dictionary could have included फ़ीका against three concepts implying that फ़ीका has three senses. As long as Hindi is the pivot, the finer concepts found in Marathi *(e.g.,* अगोड '*less sweet'*, अळणी '*less salty*, and मिळमिळत '*less spicy')* cannot be mapped to the coarse concept found in Hindi *(e.g.,* फ़ीका '*the food prepared with less sugar, salt or spice')*. However, at a later phase of the dictionary development process, the finer concepts of Marathi (or any other languages) can be identified, and added to the pivot language, *i.e.*, *Hindi*, after which the other languages can borrow the concepts from the same pivot to enrich their dictionary in the multilingual setting. The computational tool (*cf. section 8*) provides a support to mark such cases for review, and to retrieve all those when one decides to add to the pivot language synsets.

8 Computational Framework for the Multilingual Dictionary

For effective implementation of our idea of synset based multilingual dictionary, we carefully designed the dictionary development process, which is, in fact, expected to involve a number of human lexicographers. Figure 2 shows the complete semi-automatic data flow in the dictionary development process.

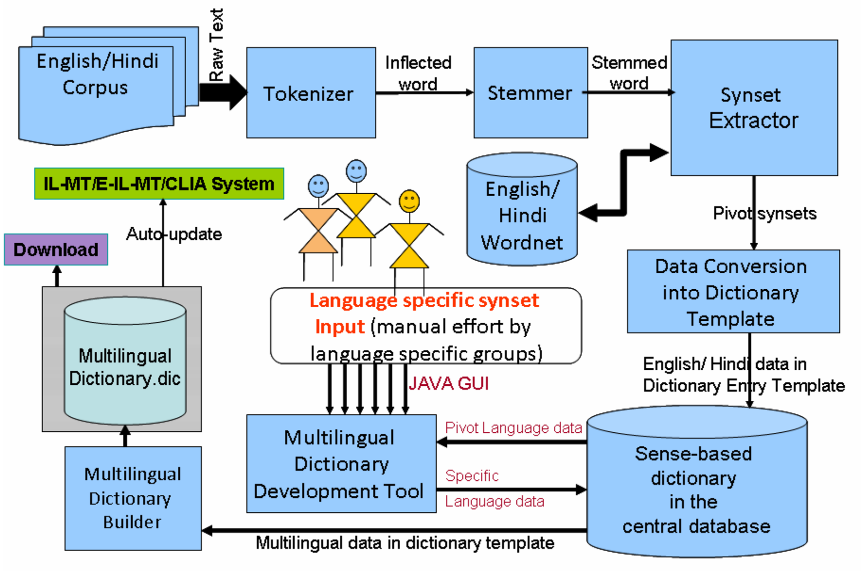

English/Hindi Corpus
Raw Text
Tokenizer
Inflected word
Stemmer
Stemmed word
Synset Extractor
English/ Hindi Wordnet
Pivot synsets
Data Conversion into Dictionary Template
English/ Hindi data in Dictionary Entry Template
IL-MT/E-IL-MT/CLIA System
Download
Auto-update
Multilingual Dictionary.dic
Multilingual Dictionary Builder
Language specific synset Input (manual effort by language specific groups)
JAVA GUI
Multilingual Dictionary Development Tool
Pivot Language data
Specific Language data
Sense-based dictionary in the central database
Multilingual data in dictionary template

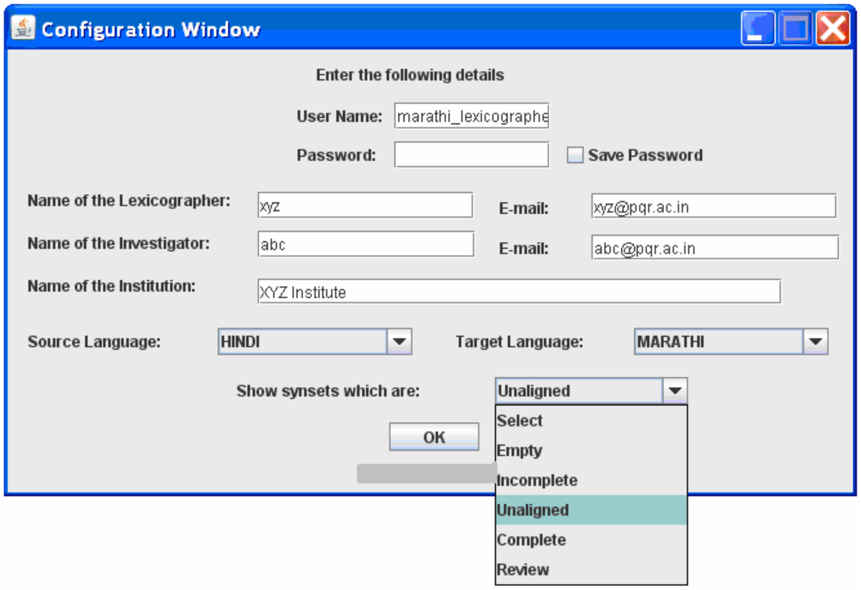
Fig. 2. Data flow in the dictionary development process

The pivot synsets are extracted from the existing Hindi wordnet along with the concept descriptions, syntactic category and examples. For convenience, an appropriate template is used for multilingual dictionary development, as illustrated in Table 3.

Table 3. Dictionary entry template

ID	:: 02691516
CAT	:: verb
CONCEPT	:: be in a state of movement or action
EXAMPLE	:: "The room abounded with screaming children"
SYNSET-ENGLISH	:: (abound, burst, bristle)

The whole process, shown in figure 2, is implemented using a centralized MYSQL database and a JAVA GUI. The screenshots of the GUI windows are shown in figure 3 and 4. Language and task configuration window is shown Figure 3, and the synset entry interface is shown in figure 4. The tool accepts the data in UNICODE only.



User Name: marathi_lexicographe

Password:

□ Save Password

Name of the Lexicographer: xyz

E-mail: xyz@pqr.ac.in

Name of the Investigator: abc

E-mail: abc@pqr.ac.in

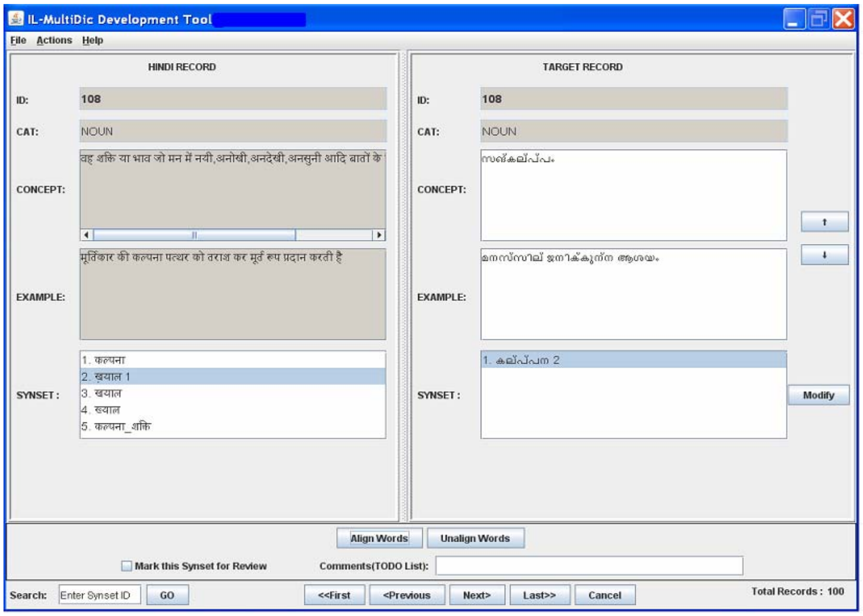
Name of the Institution: XYZ Institute

Source Language: HINDI

Target Language: MARATHI

Show synsets which are: Unaligned

Fig. 3. Language and Task Configuration Window



HINDI RECORD

ID: 108

CAT: NOUN

CONCEPT: वह शक्ति या भाव जो मन में नयी,अनोखी,अनदेखी,अनसुनी आदि बातों के

EXAMPLE: मूर्तिकार की कल्पना पत्थर को तराश कर मूर्त रूप प्रदान करती है

SYNSET :
1. कल्पना
2. ख़याल 1
3. खयाल
4. ख्याल
5. कल्पना_शक्ति

TARGET RECORD

ID: 108

CAT: NOUN

CONCEPT: സങ്കല്പ്പം.

EXAMPLE: മനസ്സില് ജനിക്കുന്ന ആശയം.

SYNSET :
1. കല്പ്പന 2

□ Mark this Synset for Review

Comments(TODO List):

Search: Enter Synset ID

Total Records : 100

Fig. 4. Synset entry and word-alignment interface

Once the dictionary is built out of the multilingual data as shown in figure 4, a lexical transfer engine provides the following for various usages:

(i) Given a word in any language, get all the records in the specified template in the same language or in any other language. *(useful for a WSD system)*

(ii) Given a word in any language and its part-of-speech, get all the records in specified template in the same language or in any other language. *(useful for a WSD system)*

(iii) Given a word in any language with respect to a particular concept, get the most appropriate translation of that word in any other language. *(useful for lexical transfer in an MT system, if a WSD system is embedded in the MT system)*

(iv) Given a word in any language, get the most probable translation of that word in any other language. *(useful for lexical transfer in an MT system having no WSD system embedded and in a cross-lingual information retrieval system)*

Using this lexical transfer engine, the multilingual dictionary is accessible online through a user-friendly website having a facility for obtaining feedback from online dictionary users. The feedback obtained from online users is expected to be useful for further development of this invaluable lexical resource.

9 Conclusion and Future Directions

We have reported here our experiences in the construction of a multilingual dictionary framework that is being used across language groups to create large scale MT and CLIR systems. Many challenges are faced on the way, chief amongst them being the one-on-one production of a target language lexeme corresponding to a source language lexeme. On the computational front there are challenges to be tackled for the maintenance of multilingual data, their insertion, deletion and updation in a spatially and temporally distributed situation. Of the many advantages of the framework are: *(i)* a linguistically sound basis of the dictionary framework, *(ii)* economy of representation and *(iii)* avoidance of duplication of effort. Our future work consists in incorporating domain sensitivity to the framework and also in solving the challenges of the distributed access and storage.

References

1. Vossen, Piek (ed.) 1999. EuroWordNet: A Multilingual Database with Lexical Semantic Networks for European languages. Kluwer Academic Publishers, Dordrecht.
2. Christodoulakis, Dimitris N. 2002 . BalkaNet: A Multilingual Semantic Network for Balkan Languages. EUROPRIX Summer School, Salzburg Austria, September 2002.
3. Miller G., R. Beckwith, C. Fellbaum, D. Gross, K.J. Miller. 1990. “Introduction to

WordNet: An On-line Lexical Database". International Journal of Lexicography, Vol 3, No.4, 235-244.

4. Fellbaum, C. (ed.) 1998, WordNet: An Electronic Lexical Database. The MIT Press.
5. Wahlster, W. (ed.). 2000. Verbmobil: Foundations of Speech-to-Speech Translation. Springer-Verlag. Berlin, Heidelberg, New York, 2000
6. Marathi Wordnet. http://www.cfilt.iitb.ac.in/wordnet/webmwn
7. Jha., S., D. Narayan, P. Pande, P. Bhattacharyya. 2001. A WordNet for Hindi. Workshop on Lexical Resources in Natural Language Processing, Hyderabad, India, January, 2001.
8. Ramanand, J., Akshay Ure, Brahm Kiran Singh and Pushpak Bhattacharyya. Mapping and Structural Analysis of Multilingual Wordnets. IEEE Data Engineering Bulletin, 30(1), March 2007.
9. Sinha, Manish., Mahesh Reddy and Pushpak Bhattacharyya. 2006. An Approach towards Construction and Application of Multilingual Indo-WordNet. 3rd Global Wordnet Conference (GWC 06), Jeju Island, Korea, January, 2006.