

BanglaNet: Towards a WordNet for Bengali Language

K.M. Tahsin Hassan Rahit

American International
University -Bangladesh
tahsin.rahit@gmail.com
Md. Al- Amin
American International
University -Bangladesh
alamin@aiub.edu

Khandaker Tabin Hasan

American International
University -Bangladesh
tabin@aiub.edu
Zahiduddin Ahmed
American International
University -Bangladesh
zahid@aiub.edu

Abstract

Despite being a popular language in the world, the Bengali language lacks in having a good wordnet. This restricts us to do NLP related research work in Bengali. Most of the today's wordnets are developed by following *expand approach*. One of the key challenges of this approach is the cross-lingual word-sense disambiguation. In our research work, we make semantic relation between Bengali wordnet and Princeton WordNet based on well-established research work in other languages. The algorithm will derive relations between concepts as well. One of our key objectives is to provide a panel for lexicographers so that they can validate and contribute to the wordnet.

1 Introduction

The Princeton WordNet (PWN) (Miller, 1995; Fellbaum, 1998) is one of the most semantically rich English lexical database which is widely used as a resource in many research and development. It is not only an important resource for NLP applications in each language, but also for inter-linking WordNets of different languages to develop multilingual applications to overcome the language barrier. In the

present, there are roughly 6,500 languages ¹. Among those, Bengali is the 7th most popular language ² in the world. Yet, there is a lack of work for Bengali wordnet. Global WordNet Association (GWA) has enlisted almost all wordnets in several levels depending on availability and how rich it is. At first level, there are 34 Open Multi-lingual WordNet ³ that are merged into Global WordNet Grid. But in spite of being a popular language, Bengali is not one of them. GWA also enlist other available wordnets. Among those 80 wordnets, there are two Bengali wordnets which are developed in India.

In this research work, a baseline for BanglaNet has been developed which is a wordnet for the Bengali language. To link the wordnet with Princeton WordNet, semi-automatic cross-lingual sense mapping approach is used. We align the Princeton WordNet synset into a bilingual dictionary through the English equivalent and its part-of-speech (POS). Manual translation and link-up can also be employed after the alignment. This paper covers previous works for other wordnets including previous

¹ How many spoken languages are there in the world, <http://www.infoplease.com/askeds/many-spoken-languages.html> (Accessed 2016-10-22)

² Most widely spoken languages in the world, <http://www.infoplease.com/ipa/A0775272.html> (Accessed 2016-10-22)

³ Open Multilingual WordNet, <http://compling.hss.ntu.edu.sg/omw/> (Accessed 23-10-2016)

attempts of developing Bengali WordNet, describe initiative taken for BanglaNet and our design and execution process in depth. Lastly, analysis of resultant lexical database has been presented. We aim to include BanglaNet into GlobalWordNet in future. Intending to doing so, relation with Princeton WordNet is maintained as much as possible as per the convention. Additionally, a web-based collaborative tool, called *Oikotan* which is BanglaNet Lexicography Development Panel (LDP) has been developed for revising the result of synset assignment and provide a framework to create BanglaNet via the linkage with synsets.

2 Background Study

2.1 WordNet Development Techniques

To this date, there are two ways develop wordnet for a particular language.

Merge Approach is used to build the word net from scratch. The Princeton WordNet is built in this approach. The taxonomies of the language, synsets, relations among synsets are developed first. Experienced work power, lexicographer and time are needed to develop for this approach (Taghizadeh and Faili, 2016). Mapping resultant wordnet with the Princeton WordNet is also required extensive work and cross-language expert.

Expand Approach is used to map or translate local words directly to the Princeton WordNet's synsets by using the existing bilingual dictionaries. Most of the WordNet available currently is developed by following this approach. This process can be made easy by semi-automatically doing many tasks and then refactoring it for further proofing.

2.2 Related Works

2.2.1 International Languages

The first attempt for developing WordNet in another language other than English started

in 1996. EuroWordNet (Vossen, 2002) began as an EU project, with the goal of developing wordnets for Dutch, Spanish and Italian and linking these wordnets to the English WordNet in a multilingual database. Later in 1997, it was extended and German, French, Czech and Estonian included. Balkan WordNet (Tufis et al., 2004) - which was developed in the BalkaNet project was developed with an aim to develop a multilingual semantic network for Balkan languages such as reek, Turkish, Romanian, Bulgarian, Czech and Serbian. In developing BalkaNet semantic relations are classified in the independent WordNets according to a shared ontology. BalkaNet was integrated along with EuroWordNet through a WordNet Management System. Relations among synsets have been built mostly automatically (Pala and Smrz, 2004) and these relations are developed based on Princeton WordNet. However, to achieve high accuracy rate developer needs to pay special attention to the problem of the translation equivalents.

There are open challenges in NLP research to automate development of semantic resources constitutes. In WOLF (Wordnet Libre du Français, Free French Wordnet) (Apidianaki and Sagot, 2012) development, multiple NLP algorithms including cross-lingual word sense disambiguation is used. WOLF is free wordnet for the French language. In Asian region, Japanese WordNet (Isahara et al., 2008) was developed using *expand approach*. Korean WordNet (Lee et al., 2002) was developed using extracting semantic hierarchy by utilizing a monolingual MRD and an existing thesaurus in *expand approach*. Thai WordNet was (Sathapornrungskij and Pluem-pitiwiriwawej, 2005) also developed by following this same approach. Another large work in Asian region includes IndoWordNet (Prabhu et al., 2012) developed in India to incorporate language used in Indian sub-continent. In-

doWordNet was also developed using existing WordNets.

Word-Sense Disambiguation (WSD) technique played a major role in most of the wordnet development. Lefever, Els and Hoste, Veronique have presented review on cross-lingual disambiguation (Lefever and Hoste, 2010) (Lefever and Hoste, 2013). They found out that languages where the ratio of word against sense is low, it becomes hard to extract translation for that language since the number of translation for a particular word in another language becomes greater. Hence, a particular word contains multiple translations in counter language.

French encountered the similar problem like us. It had no corpus with predicate-argument annotations which help to express semantic relation build-up. Van der Plas et al. researched on predicate labeling in French (van der Plas and Apidianaki, 2014) to overcome this issue using Word Sense Disambiguation.

There are two terms in cross-lingual WSD. One is *best* match and another one is *Out-of-five*. In *best* mode, the word or sense with the best probability score tagged with its counter word or sense. In case of, *Out-of-five* approach, if multiple senses or word belongs to candidate conceptualization, best five probability candidates are considered for further analysis. Further analysis can be done manually or automatically. It can be semi-automatic as well.

WSD process performance can be improved by using the Direct Semantic Transfer (DST) technique developed by Van der Plas et al. (Van der Plas et al., 2011). It tells us that the senses which can be directly transferred to another language if and only if both share same semantic property.

Surtani et al. developed a system where it can predict the paraphrases based on corpus (Surtani et al., 2013). In their system, they have a semantic relation prediction model.

Recently, BabelNet⁴ (Navigli and Ponzetto, 2012a) has become a good example of multilingual language resource. BabelNet simplified WSD process by incorporating coding API (Navigli and Ponzetto, 2012b). Primarily, it uses open-source resources such as Wikipedia. However, BabelNet does not create any WordNet for a particular language. It is a huge standalone network of multilingual resources which utilizes Princeton WordNet along with other resources to make relations.

2.2.2 Bengali

Between two of Bengali wordnets listed in GWA, one is developed by Indian Institute of Statistics under Indradhanush Project⁵. It has an online browser which does not provide the semantic relation between synsets and only provides different concept available for a word. Another Bengali wordnet is developed as part of IndoWordNet by Center for Indian Language Technology (CILT) and Indian Institute of Technology (IIT-Bombay) (Prabhu et al., 2012). A notable point in this WordNet is - it is built by following the *expand approach*. It does have the semantic relation between synset to some extent. This is the most mature and contextually rich Bengali WordNet to this date. Both WordNets are browsable and closed source. These are neither publicly available for development, use or extend nor it provides any API for general use.

There was an effort for developing Bengali WordNet in BRAC University's Center for Research on Bengali Language Processing. In their development process they followed *merge approach* (Faruqe and Khan, 2010).

⁴BabelNet can be found on <http://babelnet.org> (Accessed 2016-12-07)

⁵Indradhanush Project, <http://indradhanush.unigoa.ac.in> (Accessed 2016-10-22.)

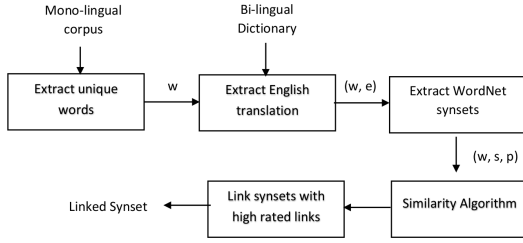


Figure 1: Proposed method for BanglaNet

3 Architecture

It has been discussed above that *expand approach* is followed to construct the BanglaNet by translating the synsets in the Princeton WordNet to the Bengali language. Both automatic and manual methods are applied in the process. Ambiguity is one of the concerns for automatic concept mapping. This cross-lingual ambiguity can come in different form. For instance - one-to-one, one-to-many, many-to-one, many-to-many. In this research work, uni-directional ambiguity in one-to-one and one-to-many has been addressed.

Based on our research on other languages’ WordNet and past works in Bengali WordNet, this paper proposes to follow methodology described in Fig 1 for BanglaNet development.

- i) Extract monosemous literals w from Bengali lexicon.
- ii) Translate each Bengali literal to English literals e using bilingual dictionary.
- iii) For each English literals, extract concept(s) available in Princeton WordNet p .
- iv) Run similarity score calculation algorithm using the e and p we found for two different Bengali sense. We take different synset available for sense w and compare their English counterpart.
- v) Based on similarity score, map Bengali concept with pwn concept.

- vi) Lexicographer validation for resultant mapping.

3.1 Similarity Matrices

In step iv, similarity algorithm is used. Similarity algorithm calculates similarity in a sense between two words in Princeton WordNet. Similarity can be calculated in several ways. There are well-established algorithms (Pedersen et al., 2004; Meng et al., 2013) to calculate similarity score. Few of those algorithms are -

- i) Path Similarity (Meng et al., 2013) calculates the score in a range of 0 to 1 based on the shortest path that connects the senses in “is-a” (hypernym/hyponym) relation.
- ii) Leacock-Chodorow Similarity (Bruce and Wiebe, 1994) scores based on the shortest path that connects the senses (identical to Path Similarity) and the maximum depth of the taxonomy in which the senses occur.
- iii) Wu-Palmer Similarity (Wu and Palmer, 1994) uses depth of the two senses in the taxonomy considering their most specific ancestor node are used to calculate the score.

There are other algorithms like Resnik Similarity (Resnik, 1995), Jiang-Conrath Similarity (Jiang and Conrath, 1997), Lin Similarity (Lin, 1998). To calculate the similarity between two concepts, we use Wu & Palmer’s similarity algorithm as it takes the hierarchical position of concepts C_1 and C_2 in the taxonomy tree relatively to the position of the most specific common concept $Iso(c_1, c_2)$ into account. It assumes that the similarity between two concepts is the function of path length and depth in path-based measures (Wu and Palmer, 1994).

$$sim_{WP}(C_1, C_2) = \frac{2 * \text{depth}(Iso(c_1, c_2))}{len(c_1, c_2) + 2 * \text{depth}(Iso(c_1, c_2))} \quad (1)$$

4 BanglaNet Development

The primary task for WordNet development using *expand approach* is to generate base lexicons and concepts. Full system including the database of Princeton WordNet is downloadable from its official website. It is possible only to get the database files without the system as well. Lexical database files can be downloaded separately as well. For base concepts, a dataset which is available on GitHub⁶ has been used. It provides conceptual gloss in Bengali for words along with its synonymy. This dataset made our work more focused on cross-lingual mapping rather than local synset construction. This research work is focused more on making relation with PWN concept rather than producing concepts. After analyzing the list of concept retrieved from the dataset, at first synsets for each concept is generated. A concept can be represented using multiple words; it ensures that we have synonyms for every concept.

Moreover, There is a POS tag available for each concept representing the word.

4.1 Word to Word Translation

Currently, a list of concepts with its gloss and synset is available. Now, English translation for each word needed to be determined. A word in one language can be represented by multiple words in another language. This is true for concept also. But for now, English translation for the enlisted words is needed. Nevertheless, for a Bengali word, there can be multiple English meaning. For example: "বল" means 'Ball' in English. It means 'Force' as well. A bilingual dictionary is needed to collect these translations. In this step, candidate translations from Bengali to English bilingual dictionary is stored. The reason behind collecting English translation using a dictionary is to

⁶Bengali Synsets Data available on GitHub, Soumenganguly. <https://github.com/soumenganguly/Bangla-Wordnet/> (Accessed 2016-10-22)

get the proper concept from WordNet. This is achieved through the WordNet concept selection algorithm which is explained in later part of this paper. For now, let's see how dictionary translations are processed.

At first, every possible English translation for each of the words in the lexicon is needed. This translation is achieved by iterating through each Bengali word in our lexicon. Bi-lingual (Bengali to English) dictionaries are used to get translations of each of the words. This translation can be from multiple parts of speech. POS for this translation is considered as well so that it can be used to properly identify correct translation in later steps. However, not all words have its counter English words. These words can be a concept which is only available in Bengali concept only. These words can also be a proper noun. For instance, the name of the places, location, river or person, scientific terms. Although, it is also possible to collect this information in run-time, to reduce time latency and run-time processing, translations along with the POS are temporarily stored.

4.2 Linking with Princeton WordNet using Probabilistic Model

It is mentioned earlier that, automated and semi-automated WordNet mostly depends on well-crafted algorithms of Natural Language Processing (NLP) and data processing. These statistical and probabilistic heuristic algorithms are good enough to create the relation between words, sense. It is obvious that the results are not always 100% accurate. Hence, lexical post-verification steps then come in place to fine tune the results.

After having the candidate translation, now it is possible to calculate the score of the probable concept from Princeton WordNet for a BanglaNet concept. Let's assume, S_c is the

synset for a Bengali concept c . We have a set of candidate translation CT_w for a particular Bengali word w . w belongs to the concept c . POS tag associated with w is a .

$$S_c = \{s \mid s \in \text{Bengali word}\} \quad (2)$$

Now, translation for each Bengali word s_i in S_c is taken:

$$ST_{s_i} = \{st_i \mid s_i \in S_c, st_i \in CT_{s_i}\} \quad (3)$$

Combining ST_{s_i} for all S_c .

$$ST_c = \{st \mid \forall st \in \bigcup_{i=0}^n ST_{s_i} \rightarrow s_i \in S_c\} \quad (4)$$

According to set theory, ST_c will contain all unique English translations for the words in Synset S_c . Synset from Princeton WordNet for each words in the set CT_w and ST_c is retrieved. POS tag for the synsets should match with a . Assuming, u as an English word -

$$syn_{u,a} = \{x \mid x \in \text{PWN Synset for } u \text{ and } x \in a\} \quad (5)$$

$$P_1 = \{x \mid \forall x \in \bigcup_{u=CT_w} syn_{u,a}\} \quad (6)$$

$$P_2 = \{x \mid \forall x \in \bigcup_{u=ST_c} syn_{u,a}\} \quad (7)$$

We take cross product of elements of P_1 against each elements of P_2 .

$$P = \{(m,n) \mid m \in CT_w \text{ and } n \in ST_c\} \quad (8)$$

After having the cross product, a similarity algorithm on each tuple is run. To calculate similarity score, equation (1) on each tuple is used. Sorting the synset P_1 according to the summation of each synset's score which is probability score for the synset, the tuple with maximum similarity score is chosen. Algorithm for this task is transcribed in Algorithm 4.1 Now, the probability score for all probable synset in Princeton WordNet for the Bengali concept is c . Bengali synset is linked with Princeton WordNet synset using

Algorithm 4.1: Algorithm for calculating probability score

```

1 Function CalculateProbabilityScore ( $P$ )
   Input:  $P$ 
   Output: Sprted scores of  $P$  based on
           probability score
2  $scores[] := \emptyset;$ 
3 foreach ( $m,n$ )  $\in P$  do
4   if  $scores[m] \neq \emptyset$  then
5      $scores[m] \leftarrow$ 
6        $scores[m] + sim_{wp}(m,n);$ 
7   else
8      $scores[m] \leftarrow sim_{wp}(m,n);$ 
9   end
10 return  $sort(scores);$ 

```

algorithm 4.2. To link Bengali concept with Princeton WordNet, multiple procedures have used to ensure correctness as much as possible. First of all, Princeton WordNet concept is assigned to those concepts in BanglaNet which have only one possible item in P_1 . Secondly, if and only if there is only one concept available for the word w , in that case, the concept from Princeton WordNet which scored high probability in probability calculation algorithm would be chosen. A point to be noted is, if any of the synonyms (word) in synset of a concept has only one concept tagged to it, it can be linked using this method. By using this first pass on all over the concepts, Princeton WordNet concepts is assigned.

5 Results and Analysis

In the initial dataset, there were 27239 unique concepts. These concepts are represented using 40158 unique words tagged with different parts of speech. Table 1 shows statistics of our initial data. In total, almost 65% of the whole concepts are tagged with Noun parts of speech.

English translation for 13029 words has

Algorithm 4.2: Algorithm for linking concept- first pass

```

1 Function LinkSynset (w)
   Input: w
2   concept count := number of concepts
   for the word w;
3   P := Generate synset cross product ;
4   sorted_scores[] :=
   CalculateProbabilityScore(P);
5   if length of sorted_scores = 1 or
   concept_count = 1 then
6     C := concepts for the word w;
7     foreach c ∈ C do
8       c.pwn ←
       sorted_scores.top().key();
9     end
10  end

```

		Noun	Adj	Verb	Adv	Total
Initial	synsets	18311	5713	2777	438	27239
	words	28311	8136	2923	788	40158
Linked	synsets	3174	1352	73	66	4665
	words	7477	2971	130	170	10748

Table 1: Status of linked Synset and Words from initial dataset

been retrieved. After applying concept linking, 4665 concepts are linked with Princeton WordNet. In total, 10748 words are linked with Princeton WordNet.

To link this 4665 concepts with Princeton WordNet, 3729 Princeton WordNet concepts are used. That means, there are cross multiple concepts within two WordNet.

Cross-lingual word-sense disambiguation can be shown using another example. For the word "ফলকপি" there are two concept available in Bengali. In English it has two concepts too.

cauliflower.n.01 a plant having a large edible head of crowded white flower buds

cauliflower.n.02 compact head of undevel-

oped white flowers

The algorithm predicted both English concepts for the two concepts available. For ফলকপি .n.01 probability score for English concepts are 4.4419589754 and 4.4419589754 respectively. On the other hand, ফলকপি .n.02 score is 6.84959684439 and 6.20774295822. It is observed that for both cases these scores are too close to prioritize probability.

Although the algorithm used in BanglaNet is directed from Bengali to English synset matching, this development can also be implied from another way around. In that case, Bengali word which represents a particular concept in Princeton WordNet can be used to verify and add more confidence to concept linking. As a result, more link up can be achieved.

Our initial synset contains gloss. But our approach does not take gloss into consideration. As a consequence, BanglaNet can be expanded using the same approach in future even if gloss for a synset is not available.

5.1 Future Works

There is a big opportunity to work on BanglaNet expansion and development. In this algorithm, the gloss is not taken into consideration. The accuracy of the algorithm can be noticeably improved by incorporating the gloss. However, a bilingual corpus will be required to achieve this. It has been found out that there is a lack of good corpus for Bengali. Good corpus is one of the key components of Natural Language Processing. However, our literature review discussed BabelNet. It's data sources and approach can be useful to map concepts. In this research work, first pass or first level linking is done. In the second pass, new algorithm needed to connect concepts which have multiple synsets in either end (BanglaNet or Princeton WordNet). We propose to use, Variable Neighborhood Search (VNS) ("Hansen and Mladenović, Nenad and MorenoA Pérez,

José A.”, 2010).

6 Conclusion

Developing wordnet is an immense task. It is our distinct pleasure that in this research work, a basic layer of the system has been laid down for Bengali wordnet from where further development can be made. Suggestion generation task for validation can be achievable through the result of this research work. Our result analysis shows that around 5000 words from initially collected data are automatically linked up with Princeton WordNet. Although there is a long way to go in the development of Bengali wordnet, this research work is starting stage for further development.

Acknowledgments

This research is funded by ICT Division - Government of the People’s Republic of Bangladesh under its fellowship program.

References

- [Apidianaki and Sagot2012] Marianna Apidianaki and Benoît Sagot. 2012. Applying cross-lingual wsd to wordnet development. In *IREC 2012-Eighth International Conference on Language Resources and Evaluation*.
- [Bruce and Wiebe1994] Rebecca Bruce and Janyce Wiebe. 1994. A new approach to word sense disambiguation. In *Proceedings of the Workshop on Human Language Technology, HLT 94*, pages 244–249.
- [Faruqe and Khan2010] Farhana Faruqe and Munit Khan. 2010. Bwn-a software platform for developing bengali wordnet. In *Innovations and Advances in Computer Sciences and Engineering*, pages 337–342. Springer.
- [Fellbaum1998] Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.
- [Hansen and Mladenović, Nenad and MorenoA Pérez, José A.”2010] Pierre Hansen and Mladenović, Nenad and MorenoA Pérez, José A.”. 2010”. “variable neighbourhood search: methods and applications”. *Annals of Operations Research*, ”175”(”1”):”367–407”.
- [Isahara et al.2008] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the japanese wordnet. In *IREC*.
- [Jiang and Conrath1997] Jay J. Jiang and David W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *CoRR*, cmp-lg/9709008.
- [Lee et al.2002] Juho Lee, Koanghi Un, Hee-Sook Bae, and Key-Sun Choi. 2002. A korean noun semantic hierarchy (wordnet) construction. In *GLOBAL WORDNET CONFERENCE*.
- [Lefever and Hoste2010] Els Lefever and Veronique Hoste. 2010. Semeval-2010 task 3: Cross-lingual word sense disambiguation. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 15–20. Association for Computational Linguistics.
- [Lefever and Hoste2013] Els Lefever and Veronique Hoste. 2013. Semeval-2013 task 10: Cross-lingual word sense disambiguation. *Proc. of SemEval*, pages 158–166.
- [Lin1998] Dekang Lin. 1998. An information-theoretic definition of similarity. *Proceedings of ICML*, pages 296–304.
- [Meng et al.2013] Lingling Meng, Runqing Huang, and Junzhong Gu. 2013. A review of semantic similarity measures in wordnet. *International Journal of Hybrid Information Technology*, 6(1):1–12.
- [Miller1995] George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- [Navigli and Ponzetto2012a] Roberto Navigli and Simone Paolo Ponzetto. 2012a. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

- [Navigli and Ponzetto2012b] Roberto Navigli and Simone Paolo Ponzetto. 2012b. Multilingual wsd with just a few lines of code: the babelnet api. In *Proceedings of the ACL 2012 System Demonstrations*, pages 67–72. Association for Computational Linguistics.
- [Pala and Smrz2004] Karel Pala and Pavel Smrz. 2004. Building czech wordnet. *Romanian Journal of Information Science and Technology*, 7(2-3):79–88.
- [Pedersen et al.2004] T Pedersen, S Patwardhan, and J Michelizzi. 2004. Wordnet similarity - measuring the relatedness of concepts. *Proceedings - Nineteenth National Conference on Artificial Intelligence (AAAI-2004): Sixteenth Innovative Applications of Artificial Intelligence Conference (IAAI-2004)*, pages 1024–1025.
- [Prabhu et al.2012] Venkatesh Prabhu, Shilpa Desai, Hanumant Redkar, NR Prabhugaonkar, Apurva Nagvenkar, and Ramdas Karmali. 2012. An efficient database design for indowordnet development using hybrid approach. *3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP)*.
- [Resnik1995] Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'95*, pages 448–453, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- [Sathapornrungskij and Pluempitiwiriawej2005] Patanakul Sathapornrungskij and Charnyote Pluempitiwiriawej. 2005. Construction of thai wordnet lexical database from machine readable dictionaries. *Proc. 10th Machine Translation Summit, Phuket, Thailand*.
- [Surtani et al.2013] Nitesh Surtani, Arpita Batra, Urmi Ghosh, and Soma Paul. 2013. Iiith: A corpus-driven co-occurrence based probabilistic model for noun compound paraphrasing. *Atlanta, Georgia, USA*, page 153.
- [Taghizadeh and Faili2016] Nasrin Taghizadeh and Hesham Faili. 2016. Automatic wordnet development for low-resource languages using cross-lingual wsd. *Journal of Artificial Intelligence Research*, 56:61–87.
- [Tufis et al.2004] Dan Tufis, Dan Cristea, and Sofia Stamou. 2004. Balkanet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information science and technology*, 7(1-2):9–43.
- [van der Plas and Apidianaki2014] Lonneke van der Plas and Marianna Apidianaki. 2014. Cross-lingual word sense disambiguation for predicate labelling of french. In *Proceedings of the 21st TALN (Traitement Automatique des Langues Naturelles) conference*, pages 46–55.
- [Van der Plas et al.2011] Lonneke Van der Plas, Paola Merlo, and James Henderson. 2011. Scaling up automatic cross-lingual semantic role annotation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 299–304. Association for Computational Linguistics.
- [Vossen2002] Piek Vossen. 2002. Wordnet, eurowordnet and global wordnet. *Revue française de linguistique appliquée*, 7(1):27–38.
- [Wu and Palmer1994] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics, ACL '94*, pages 133–138, Stroudsburg, PA, USA. Association for Computational Linguistics.