Abstract—This paper reports the process of developing WordNet Affect lists in Bengali from the affect wordlists already available in English. It is organized in six basic emotions such as anger, disgust, fear, joy, sadness and surprise. Emotion or affect words of the six lists are updated using equivalent synsets of the SentiWordNet by keeping their parts-of-speech information unchanged. Another updating process employed especially for the verb entries is carried out on six affect lists with the help of VerbNet. The updated lists are converted into Bengali using the synset based English to Bengali bilingual dictionary. The sense disambiguation task is conducted based on the hints of sense wise separated word groups present in Bengali to English bilingual dictionary. Human translator translates the non-translated emotion word entries into Bengali. The statistical inter-translator agreement between two translators is measured using kappa coefficient (k) that shows a moderate agreement varying from 0.44 to 0.56 for six emotion classes.

Keywords- WordNet Affect; SentiWordNet; VerbNet; Bilingual Dictionary; Translation; Sense Disambiguation; Kappa

I. INTRODUCTION

Affect analysis is a natural language processing (NLP) technique for recognizing the emotive aspect of text. The same textual content can be presented with different emotional slants [3]. Emotion analysis, a recent sub discipline at the crossroads of information retrieval [18] and computational linguistics [13] is also becoming increasingly important as an application of affective computing. Human emotion described in texts is an important cue for our daily communication. But, the identification of emotional state from texts is not an easy task as emotion is not open to any objective observation or verification [17]. The majority of subjective analysis methods that are related to opinion or emotion is based on textual keywords spotting and therefore explores the necessity to build specific lexical resources.

Affective lexicon [7] is one of the most efficient resources. SentiWordNet, [3] used in opinion mining and sentiment analysis assigns to each synset of WordNet [2] three sentiment scores such as positive, negative and objective. Subjectivity wordlist [9] assigns words with the strong or weak subjectivity and prior polarities of types positive, negative and neutral. WordNet Affect [7] contains words that convey emotion. It is a small well-used lexical resource but valuable for its affective annotation. To the best of our knowledge, all of these lexical resources have been created for English. But, the amount of the text data written in languages other than English are rapidly increasing [10]. Recent study shows that non-native English speakers support the growing use of the Internet. This raises the demand for automatic text analysis tools and linguistic resources for languages other than English. Bengali is the fifth popular language in the World, second in India and the national language in Bangladesh but it is less privileged and less computerized compared to English. Resource acquisition is one of the most challenging obstacles to work with resource constrained language Bengali. Although the works on emotion analysis in Bengali have started recently [11] but there is no existing full-fledged emotion lexicon in Bengali yet. On the other hand, blogs, social networks, chats are becoming the communicative and informative repository of text based emotional contents in the Web 2.0 with a booming growth. The enormous volume of texts with emotionally rich content grows in geometrical progression and therefore makes the task of affective text analysis more crucial.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the same VerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the same VerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the sameVerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the same VerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the same VerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.

The collection of the WordNet Affect synsets used in the present task was provided as a resource for the SemEval-2007 shared task of “Affective Text”. The shared task was focused on text annotation by affective tags [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s six emotional category labels: joy, fear, anger, sadness, disgust and surprise. The task reported in this paper aims to develop WordNet Affect lists in Bengali from the available English WordNet Affect lists [7] without considering the problems of the lexical affect representation.

The lists are updated with the synsets retrieved from the English SentiWordNet [3] to make adequate number of emotion word entries. The part-of-speech (POS) information for each of the synsets is kept unchanged.

Member verbs present in the same VerbNet [15] class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The member verbs belonging to the same VerbNet class produce the verb synsets. The VerbNet classes are provided in XML file format. If a verb from any of the six affect lists is also present in a VerbNet synset, then the corresponding affect list is updated with the members of the VerbNet synset. A duplicate removal technique is applied on the six affect lists whenever these are updated using either the SentiWordNet or the VerbNet synsets.
The lists are automatically translated into Bengali using the synset based English to Bengali bilingual dictionary being developed as part of the EILMT\textsuperscript{b} project. The duplicate removal technique is also applied on the translated synsets. Human translator translates the non-translated entries containing word combinations, idioms etc. Sense wise separated word groups give a clue of pattern-based similarity in Bengali to English bilingual dictionary\textsuperscript{c}. The sense disambiguation algorithm based on similarity clue is applied on the translated Bengali synsets.

Two translators carry out the evaluation. Inter-translator agreement is done through a statistical measure, kappa [14]. The kappa coefficient ($k$) varies in the range from 0.44 to 0.56. It shows a moderate agreement and achieves significant impact on the overall translation process. There is no drastic difference between the two results as one of which is obtained after the sense disambiguation process and another one is achieved after inter-translator agreement containing "yes-yes" binary decision. The automatic sense disambiguation technique for reducing human effort is therefore considered as an effective contribution.

The rest of the paper is organized as follows. The updating of resources is described in Section II. The automatic translation of synsets and the automatic sense disambiguation task are specified in Section III and Section IV respectively. Section V describes the method of calculating kappa for inter-translator agreement. Finally Section VI concludes the paper.

II. UPDATING OF WORDNET AFFECT

A. WordNet Affect

The English WordNet Affect, based on Ekman’s six emotion types, is a lexical resource containing information about the emotional words. Compared to the complete WordNet [2], WordNet Affect is a smaller lexical resource but its affective annotation helps in emotion analysis. The collection of WordNet Affect synsets was provided as a resource for the shared task of Affective Text in SemEval-2007 [6]. A portion of the WordNet Affect synsets [8] were further annotated using Ekman’s [16] six emotional category labels: joy, fear, anger, sadness, disgust and surprise. WordNet Affect [7] is developed based on WordNet domains [5] where each synset is annotated with at least one domain label, selected from a set of around two hundred labels that are arranged hierarchically.

Concentrating only on the problem that how the affective meanings are expressed in the natural language, “Affective Words” bears only emotional connotation [1]. Not only the words that describe specific emotions (for example, joy, sad or score) but also the emotional words describing mental states, physical or bodily states, personality traits, behaviors, attitudes, and feelings (such as pleasure or pain) are present in the WordNet Affect lists.

The underlying differences among emotions, cognitive states and affects are not analyzed in the present work. Our main focus in the task is to develop an equivalent WordNet Affect resource in Bengali labeled with six emotions. The resource is provided in six separate files named by the six Ekman’s emotions. Each of the files contains a list of synsets and one synset per line. An example synset entry from WordNet Affect is shown as follows.

\[
\text{a}\#00117872 \text{ angered enraged furious infuriated maddened}
\]

The first letter of each line indicates the part of speech (POS) and is followed by the synset identification number. The representation is simple and amenable for further processing. There are a large number of word combinations, collocations and idioms in the WordNet Affect. These types of items in the synsets show problem during translation and therefore manual translation is incorporated to carry out the translation.

B. Update using SentiWordNet

It is found that the WordNet Affect lists [7] contain fewer emotion words. Hence, the lists require an update before translating them into Bengali to make adequate number of emotion words in the lists.

The WordNet Affect lists are updated with the synsets retrieved from the SentiWordNet [3], a lexical resource used for opinion mining. WordNet Affect lists as well as SentiWordNet are developed from the WordNet [2]. SentiWordNet assigns three sentiment scores such as positive, negative and objective to each synset of WordNet and contains a large number of sentiment words. One example entry of the SentiWordNet is given as follows.

\[
a\ 121184\ 0.25\ 0.25\ \text{infuriated}\#a\#1\ \text{furious}\#a\#2\ \text{maddened}\#a\#1\ \text{enraged}\#a\#1\ \text{angered}\#a\#1\n\]

Our aim is to increase the number of emotion words in the WordNet Affect lists. To accomplish the purpose, each word of the WordNet Affect synset is replaced by the equivalent synsets of SentiWordNet if the SentiWordNet synsets contain that emotion word. The part of speech (POS) information of the WordNet Affect synsets is kept unchanged.

It is also found that the equivalent synsets of SentiWordNet for some emotion words (e.g. huffiness, offense) of WordNet Affect may or may not contain a subjective score. The examples are as follows.

\[
n\#05589074 \text{ huffiness} /\text{ WordNet Affect */}
\]

\[
n\ 7057022\ 0.0\ 0.0\ \text{huffiness}\#n\#1
\]

\[
*/\text{ SentiWordNet */}
\]

\[
#05588822 \text{ umbrage} \text{ offense} /\text{ WordNet Affect */}
\]

\[
n\ 7590773\ 0.0\ 0.0\ \text{offence}\#n\#2\ \text{offense}\#n\#4
\]

\[
*/\text{ SentiWordNet */}
\]

\textsuperscript{b} English to Indian Languages Machine Translation (EILMT) is a TDIL project undertaken by the consortium of different premier institutes and sponsored by MCIT, Govt. of India

\textsuperscript{c} http://home.uchicago.edu/~cbs2/banglainstruction.html
The equivalent synsets of SentiWordNet for an individual emotion word of the WordNet Affect are added to make a new synset for that emotion word. But, a synset of WordNet Affect may contain multiple emotion words. Each of the words may or may not produce a new synset. Hence, the newly produced synsets and the non-replaced emotion words are merged to produce a single equivalent synset. Therefore the numbers of synsets in six affect lists remain same as before. The updating results of six WordNet Affect are shown in Table I.

But, each of the newly updated synsets may contain duplicate emotion words. The objective of the updation task is not only to increase the number of emotion words in the lists but to remove the number of duplicate entries also. Hence, we have applied one technique for removing the duplicates from the newly updated synsets.

For example, If the emotion words “A” and “B” in WordNet Affect synset “E” are replaced by the equivalent synsets A and B as retrieved from SentiWordNet, then the newly produced equivalent synset, E’ = (A - B’) + (B’ - A’) + (A ∩ B’). The emotion words A1, A2, A3, B3 ∈ A and B1, B2, B3, A3 ∈ B. The equivalent newly produced synset E’ contains the emotion words A1, A2, A3, B1, B2, and B3 without any duplicate. One example entry of the replaced and updated synset of WordNet Affect is shown as follows.

/* WordNet Synset */

n!10337658 fit(A) scene(B) tantrum
/* SentiWordNet Synset for A */

tantrum/scene/conniption/fit/burst/fit_out/equip/outfit/tally/jibe/match/correspond/gibe/agree/check/conform_to/meet/set/primed/fit_to/fit_for/convulsion/paroxysm

/* SentiWordNet Synset for B */

tantrum/scene/conniption/fit/scenery/view/prospect/vista/panorama/aspect/shot

/* Updated Synset E’ */

tantrum/scene/conniption/fit/burst/fit_out/equip/outfit/tally/jibe/match/correspond/gibe/agree/check/conform_to/meet/set/primed/fit_to/fit_for/convulsion/paroxysm/scenery/view/prospect/vista/panorama/aspect/shot

C. Updation using VerbNet

A special initiative is taken for updating the synsets of the emotional verbs in the WordNet Affect lists. The updating process is similar to the approach as described in the earlier section. Each emotion verb of WordNet Affect lists is replaced again using the synsets retrieved from the VerbNet.

VerbNet (VN) [15] is the largest online verb lexicon with explicitly stated syntactic and semantic information based on Levin’s verb classification [4]. VerbNet associates the semantics of a verb with its syntactic frames and combines traditional lexical semantic information such as thematic roles and semantic predicates, with syntactic frames and selectional restrictions.

<table>
<thead>
<tr>
<th>WordNet Affect Lists</th>
<th>Before Updating</th>
<th>After Updating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#Number of words (#Number of synsets)</td>
<td>#Number of words</td>
</tr>
<tr>
<td>Anger</td>
<td>318 (128)</td>
<td>544</td>
</tr>
<tr>
<td>Disgust</td>
<td>72 (20)</td>
<td>104</td>
</tr>
<tr>
<td>fear</td>
<td>208 (83)</td>
<td>371</td>
</tr>
<tr>
<td>joy</td>
<td>539 (228)</td>
<td>904</td>
</tr>
<tr>
<td>Sadness</td>
<td>309 (124)</td>
<td>309</td>
</tr>
<tr>
<td>Surprise</td>
<td>90 (29)</td>
<td>99</td>
</tr>
</tbody>
</table>

The member verbs in the same VerbNet class share common syntactic frames, and thus they are believed to have the same syntactic behavior. The VerbNet files containing the member verbs with similar sense are stored in a XML format. Hence, we have considered that the member verbs present in the same class are sense based synonymous verbs. We have prepared each of the synsets by extracting the similar sensed synonymous verbs from each of the VerbNet classes. Irrespective of other information, a general list (VerbGL) containing only the verb synsets is prepared from all of the VerbNet XML files.

Each of the emotion verbs present in the updated Wordnet Affect lists is searched in the general list (VerbGL). If any match occurs, the corresponding verb is replaced by the synsets available in the general list (VerbGL). The results regarding the updation of emotion verbs in the WordNet Affect lists are shown in Table II. The technique similar to that described in Section II.B is applied to the newly updated verb synsets for removing the duplicates.

<table>
<thead>
<tr>
<th>WordNet Affect Lists</th>
<th>Before Updating</th>
<th>After Updating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#Number of words (#Number of Verbs)</td>
<td>#Number of words (#Number of Verbs)</td>
</tr>
<tr>
<td>anger</td>
<td>544 (39)</td>
<td>765 (56)</td>
</tr>
<tr>
<td>disgust</td>
<td>104 (12)</td>
<td>195 (25)</td>
</tr>
<tr>
<td>fear</td>
<td>371 (32)</td>
<td>566 (51)</td>
</tr>
<tr>
<td>joy</td>
<td>904 (44)</td>
<td>1824 (69)</td>
</tr>
<tr>
<td>sadness</td>
<td>309 (28)</td>
<td>852 (39)</td>
</tr>
<tr>
<td>surprise</td>
<td>99 (28)</td>
<td>260 (53)</td>
</tr>
</tbody>
</table>

III. AUTOMATIC TRANSLATION

To accomplish the translation task, an English-to-Bengali bilingual synset based dictionary containing approximately 1,02,119 entries is used. The dictionary is being developed as
part of the EILMT project with the help of Samsad Bengali to English bilingual dictionary. We have used English as a source language to prepare the bilingual dictionary.

We have not considered all of the word combinations, as they could not be translated automatically. Some of the synsets of WordNet Affect lists are not automatically translated. Some words containing suffixes such as “-ness”, “-less”, “-ful” as well as some adverbs formed using suffix “-ly” are unlikely to appear in the dictionary. Although the synset-based dictionary is being developed for the general domain but some of the words are not translated as such words may not be found in the bilingual dictionary.

The non-translated entries are filtered from the English WordNet Affect after the translation process. The number of non-translated words of Bengali WordNet Affect lists is shown in Table III. The total number of non-translated words in the six emotion lists is 210 and the number is comprehensible for manual translation. The idioms and word combinations are not translated automatically and manual effort is applied for translating these non-translated items. It is found that the total number of emotion words increases in all of the WordNet Affect lists except joy and sadness. The duplicate removal technique used in Section II.B is also applied on the translated Bengali synsets as the translation process produces duplicate Bengali emotion words. The total number of translated synsets along with the manually translated words in six Bengali WordNet Affect lists is shown in Table IV. The example entry of a Bengali translated synset is shown in Figure 1.

![Figure 1. Example of a Translated Bengali Synset](image)

**TABLE III. NUMBER OF TRANSLATED AND NON-TRANSLATED EMOTION WORKS WITH NUMBER OF TRANSLATED SYNTENS**

<table>
<thead>
<tr>
<th>WordNet Affect Lists</th>
<th>#Number of translated words (#Number of synsets)</th>
<th>#Number of non-translated words</th>
</tr>
</thead>
<tbody>
<tr>
<td>anger</td>
<td>1141 (321)</td>
<td>80</td>
</tr>
<tr>
<td>disgust</td>
<td>287 (74)</td>
<td>37</td>
</tr>
<tr>
<td>fear</td>
<td>785 (182)</td>
<td>27</td>
</tr>
<tr>
<td>joy</td>
<td>1644 (467)</td>
<td>42</td>
</tr>
<tr>
<td>sadness</td>
<td>788 (220)</td>
<td>10</td>
</tr>
<tr>
<td>surprise</td>
<td>472 (125)</td>
<td>14</td>
</tr>
</tbody>
</table>

**IV. AUTOMATIC SENSE DISAMBIGUATION**

An automatic sense disambiguation technique that is similar to the approach adopted for Bengali verb subcategorization frame identification tasks (Banerjee et al., 2009) [19] is employed to the translated synsets. The disambiguation task is carried out by classifying the translated synsets into the group of synonyms containing similar sense. The technique adopted here not only aims for conducting the disambiguation task but also to minimize the manual effort necessary to remove any discrepancy in the ambiguous Bengali synsets. The ambiguity issues are resolved by conducting mutual agreement between the translators. The agreement issues in detail are discussed in the next Section.

We have introduced an algorithm for disambiguating the words belonging to same as well as different synsets automatically by keeping the total number of emotion words unchanged in six Bengali WordNet Affect lists. The pattern based informative clue present in the Bengali to English bilingual dictionary helps us to accomplish the task as well. Each of the words from a Bengali translated synset is searched in the Bengali to English bilingual dictionary to extract their English equivalent synonyms of the same and different senses. The example entries present in the Bengali to English bilingual dictionary for the emotional words 〈ক্রুদ্ধ (kruddha) and প্রকুপিত (prakupita)〉 of a translated Bengali synset are given as follows.

# Member Words:

< ক্রুদ্ধ [ kruddha ] a angry; angered, enraged, wrathful; indignant.>

< প্রকুপিত [ prakupita ] a enraged, angered, in censed, infuriated; excited.>

In the dictionary, different synonyms for a word with the same sense are separated using “,” and different senses are separated using “;”. The synonyms for the different senses of a word are extracted from the dictionary. Each group of English equivalents yields a resulting set called Synonymous Word Set (SWS).

For example, the English synonyms (angered, enraged) with the same sense and synonym with another senses (wrathful), (indignant) are retrieved for the Bengali key word ক্রুদ্ধ (kruddha) and are categorized as three different SWSs for that Bengali key word. Similarly, two SWSs are formed for the word “ঝুঞ্জা ” (rupsta). Two Bengali emotion words belong to the same or different translated synsets are grouped together to form a new Bengali synset if there is at least one common English word (e.g angered) present in both English equivalent SWSs of the corresponding Bengali words. For example, two English equivalent classes Cxb and Cyb with respect to two Bengali words Xb and Yb are defined as follows,

\[
Cxb = \{SWS1, SWS2, \ldots, SWSq\} \\
Cyb = \{SWS1, SWS2, \ldots, SWSp\}
\]

If, for \(i = 1 \leq p, j = 1 \leq q\), \((SWSi \cap SWSj) \neq \phi\), or \(\exists Ze | Ze \in SWSi \cap SWSj\), where Ze is the equivalent English word present in any of the Synonyms Word Sets (SWS) of Cxb and
TABLE IV. NUMBER OF SYNSETS AFTER SENSE DISAMBIGUATION AND NUMBER OF EMOTION WORDS PRESENT IN THE LISTS

<table>
<thead>
<tr>
<th>WordNet Affect Lists</th>
<th>After Sense Disambiguation (#Number of synsets)</th>
<th>After Manual Translation of not translated words</th>
<th>After Agreement (Y-Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>1141 (368)</td>
<td>1221</td>
<td>927</td>
</tr>
<tr>
<td>Disgust</td>
<td>287 (110)</td>
<td>324</td>
<td>274</td>
</tr>
<tr>
<td>Fear</td>
<td>785 (243)</td>
<td>812</td>
<td>630</td>
</tr>
<tr>
<td>Joy</td>
<td>1644 (507)</td>
<td>1686</td>
<td>1488</td>
</tr>
<tr>
<td>Sadness</td>
<td>788 (280)</td>
<td>798</td>
<td>610</td>
</tr>
<tr>
<td>Surprise</td>
<td>472 (189)</td>
<td>486</td>
<td>426</td>
</tr>
</tbody>
</table>

V. AGREEMENT MEASURE

In our present task, we have used the standard metric for conducting the inter-translator agreement. The agreement is carried out for the emotional words present in the translated Bengali synsets. Cohen's kappa coefficient \(k\) is adopted for conducting the agreement measure. Cohen's kappa coefficient \(k\) is a statistical measure of inter-rater agreement for qualitative (categorical) items. It measures the agreement between two raters who separately classify items into some mutually exclusive categories. The equation for \(k\) is:

\[
κ = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}
\]

where \(Pr(a)\) is the relative observed agreement among raters, and \(Pr(e)\) is the hypothetical probability of chance agreement. If the raters are in complete agreement then \(κ = 1\). If there is no agreement among the raters, then the value of \(κ\) is considered as \(≤ 0\). The inter-translator agreements for the emotion words of each of the emotion lists are shown in Table V.

TABLE V. INTER-TRANSLATOR AGREEMENT

<table>
<thead>
<tr>
<th>WordNet Affect Lists</th>
<th>Agreement Values</th>
<th>Kappa ((k))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>anger</td>
<td>927</td>
<td>110</td>
</tr>
<tr>
<td>disgust</td>
<td>274</td>
<td>16</td>
</tr>
<tr>
<td>fear</td>
<td>630</td>
<td>50</td>
</tr>
<tr>
<td>joy</td>
<td>1488</td>
<td>62</td>
</tr>
<tr>
<td>sadness</td>
<td>610</td>
<td>53</td>
</tr>
<tr>
<td>surprise</td>
<td>426</td>
<td>18</td>
</tr>
</tbody>
</table>

Two translators carry out the overall agreement to verify the presence of the translated Bengali emotion words in their respective WordNet Affect list. The translators consider only binary decision Yes or No (Y/N) for finally assigning each emotion word to its corresponding WordNet Affect list. The present work shows that the agreement values for six WordNet Affect lists are in the range from 0.44 to 0.56 that gives a significantly moderate value. It has to be mentioned that the inter translator agreement with “Yes-Yes” combination gives the number of emotion words in each of the six lists satisfactory. The results of the number of emotion word entries after the sense disambiguation process as well as after the agreement measure are shown in Table V. The small
differences in number of words of the six lists after disambiguation task and agreement measure indicate the usefulness of incorporating the automatic sense disambiguation process as the disambiguation technique reduces the manual effort significantly. The disagreement occurs mostly for the synsets containing emotion words more than five. But, the negotiation among the translators reduces clarity regarding the disagreement issues.

VI. CONCLUSION

The paper describes the preparation of Bengali WordNet Affect containing six types of emotion words in six separate lists. The automatic way of updating, translation and sense disambiguation task reduces the manual effort. The intertranslator agreement is also moderate. Although, the agreement is moderate but the most important future task is to evaluate the resource in each step of updating, translation and sense disambiguation using some standard metrics. The resource is still under development. But, a part of it can be provided on request. The lists of the Bengali WordNet Affect can be used for emotion related language processing tasks in Bengali. Our future task is to integrate more resources so that the number of emotion word entries in the lists increases. The sense disambiguation task needs to be improved further in future attempts by incorporating more number of translators and considering their agreement into account. Similar technique can be used for developing this type of resources for other languages also. Our future plan is to explore the resource by incorporating cognitive knowledge also. The effectiveness and robustness are to be investigated in the web scale (especially, in the noise data including web logs and social web where it is most useful) for future perspectives.

REFERENCES


