Identifying Emotional and Polar Concepts via Synset Translation

Logan Woudstra, Moyo Dawodu, Frances Igwe
Senyu Li, Ning Shi, Bradley Hauer, Grzegorz Kondrak
Alberta Machine Intelligence Institute (Amii)
Department of Computing Science
University of Alberta, Edmonton, Canada
{lwoudstr,mdawodu,figwe,senyu,ning.shi,bmhauer,gkondrak}@ualberta.ca

Abstract

Emotion identification and polarity classification seek to determine the sentiment expressed by a writer. Sentiment lexicons that provide classifications at the word level fail to distinguish between different senses of polysemous words. To address this problem, we propose a translation-based method for labeling each individual lexical concept and word sense. Specifically, we translate synsets into 20 different languages and verify the sentiment of these translations in multilingual sentiment lexicons. By applying our method to all WordNet synsets, we produce SentiSynset, a synset-level sentiment resource containing 12,429 emotional synsets and 15,567 polar synsets, which is significantly larger than previous resources. Experimental evaluation shows that our method outperforms prior automated methods that classify word senses, in addition to outperforming ChatGPT. We make the resulting resource publicly available on GitHub.

1 Introduction

Emotion identification is the semantic task of analyzing a piece of text to identify a set of underlying emotions from a predefined inventory (de Albornoz et al., 2012). *Polarity classification* is the closely related task of determining the polarity of a text, which can be positive, negative, or neutral (Pang and Lee, 2004; Turney, 2002). These two tasks are variations on *sentiment analysis*, the extraction of sentiment that a writer expresses toward some object (Jurafsky and Martin, 2009). Following Kakkonen and Galić Kakkonen (2011), we refer to a text, a word token, or a lexical concept as *sentimental* if it is associated with any emotion or non-neutral polarity.

The scope of sentiment analysis can be a single word (Pennebaker et al., 2001; Mohammad and Turney, 2010, 2013), a sentence (Abdul-Mageed and Ungar, 2017; Sosea and Caragea, 2020), or longer texts such as Twitter posts and customer reviews (Liew and Turtle, 2016; Dini and Bittar, 2016; Hu and Liu, 2004). In this paper, we focus on *sense-level sentiment*; knowing the sentiment of the individual words in a text can help determine its overall sentiment.

Emotion identification is more informative than polarity classification, but it is also more subjective in the sense that we would expect more disagreement among annotators. For example, determining that the word *murder* has a negative polarity is more objective than deciding which combination of emotions, such as anger, disgust, fear, and sadness, best relate to the word. This subjectivity is only heightened by the lack of consensus on the set of basic human emotions. Researchers have proposed inventories of six (Ekman, 1992), eight (Plutchik, 1962), or more fundamental emotions. Therefore, while we explore both tasks, we place greater emphasis on polarity classification.

Since many emotion-bearing words are polysemous, we focus our attention on word senses and lexical concepts. Senses are associated with one specific meaning of a word, so classifying sentiments at the level of senses avoids the ambiguity that arises from words having multiple meanings. In WordNet (Miller, 1995), sets of words that express the same concept are grouped together in *synsets*, each uniquely corresponding to a single concept. For example, the synset that contains the words *sadness*, *sorrow*, and *sorrowfulness* corresponds to the concept which is defined as "the state of being sad". Synsets in WordNet are connected via various relations. A word can convey different sentiments depending on its sense in a given context; we assume that the sentiment associated with a specific sense/synset/concept is fixed. While it is true that the sentiment of a sense can too change depending on the context in which it is used, this ambiguity is much less prevalent among senses than it is among words. Thus, by labeling a synset,

*Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024)*, pages 142 - 152
June 20-21, 2024

we provide a single emotional label for all word senses in the synset.

Furthermore, we hypothesize that the sentiment of a given concept is likely to be the same in other languages. For example, the concept mentioned above is also expressed by the Spanish word *tristeza* and the Yoruba word *ibanuje*. We test this hypothesis by developing methods that classify English word senses by leveraging multilingual translations. Conversely, we leverage English sentiment labels for other languages.

In this paper, we outline the development of an automatic method that leverages multilinguality to identify sentimental concepts. Unlike existing resources that were constructed by expanding a core of manually-annotated synsets, we propose a fully automatic method that can provide labels for a significantly larger number of synsets. Our method achieves a precision of 96.0% and 92.0% on identifying emotional and polar synsets, respectively. Of those, a correct emotional label is assigned with an accuracy of 84.3%, and a correct polarity label is assigned with an accuracy of 95.8%. The resulting resource, which we call *SentiSynset*, contains 12,429 emotional and 15,567 polar synset labels. When used in conjunction with word sense disambiguation techniques, the resource could be useful for the downstream application of sentiment analysis at the level of sentences and documents. SentiSynset is publicly available on GitHub, together with our code.[1]

2 Related Work

In this section, we provide an overview of the related work on emotion identification and polarity classification at the synset level. Our focus is on the resources based on the Princeton WordNet (Miller, 1995), which consists of 117,659 synsets, each corresponding to a specific concept defined by its *gloss*.

Emotion identification WordNet-Affect (Strapparava and Valitutti, 2004; Strapparava et al., 2006) and SentiSense (de Albornoz et al., 2012; Carrillo-de Albornoz and Plaza, 2013) associate a subset of WordNet synsets with emotional classifications. WordNet-Affect contains 2,874 synsets, each associated with one or more of 32 emotions. It was constructed by first manually annotating a relatively small "core" of emotional synsets, which was later expanded by leveraging inter-synset relations in WordNet. SentiSense encompasses 2,190 synsets labeled with one of 14 emotional categories. While its development is similar to that of WordNet-Affect, they differ in their specific sets of manually annotated synsets and the WordNet relations chosen for extension.

WordNet-Affect and SentiSense are built upon emotional inventories that are not only mutually incompatible but also rooted in separate psychological theories of emotion. This misalignment complicates data integration, consistency maintenance, and interpretation. Meanwhile, reconciling the two resources by mapping their distinct emotion inventories remains problematic. For example, senses of the words *abashed* and *upset* are both identified with *anxiety* in WordNet-Affect, but are respectively labeled with *disgust* and *anger* in SentiSense; however, senses of the words *embarrassment* and *nausea* are both identified with *disgust* in SentiSense, but are respectively labeled with *shame* and *general-dislike* in WordNet-Affect. These discrepancies highlight the inherent subjectivity in emotion identification, thus motivating our prioritization of the more objective task of polarity classification. Additionally, both resources provide limited coverage of WordNet of less than 3,000 synsets each; this limited coverage arises from their semi-automatic construction. We aim to address this problem by developing a scalable automatic method that can classify a much larger proportion of WordNet synsets.

Polarity classification SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010) stands as a prominent resource for polarity classification. It assigns each synset a positive, negative, and objective score, with values ranging from 0.0 to 1.0, summing up to 1 across the three categories. These scores are produced by a committee of classifiers which leverage synset glosses. Since the method is entirely automated, polarity scores are assigned to every WordNet synset. Contrariwise, our method, while automated, is focused on precision, rather than coverage; we do not seek to label every synset, but rather aim to label as many synsets as possible with high confidence.

ML-SentiCon (Cruz et al., 2014) attains polarity labels for synsets using a variation of the method used to create SentiWordNet. As such, the resource has the same drawbacks as SentiWordNet. In addition to the synset labels, ML-Senticon also contains

[1] https://github.com/UAlberta-NLP/SentiSynset

lemma-level lexicons for English, Spanish, Catalan, Basque, and Galician that were developed by averaging the polarity values of all synsets belonging to a lemma. While these are useful lexicons, particularly because of the inclusion of low-resource languages, assigning labels to lemmas introduces issues with polysemy.

Multilinguality Chen and Skiena (2014) leverage multilingual information to develop word-level polarity lexicons for 136 major world languages. They create graphs connecting words from these languages, considering both cross-language links, such as translations and transliterations, and intra-language links, such as synonyms and antonyms. They propagate English word-level polarity labels across the graphs to create lexicons for the non-English languages. While these automatically developed lexicons have high levels of agreement with human-annotated lexicons, they still retain the ambiguity that arises when sentiment labels are assigned to words rather than senses.

Applications of Synset Lexicons Synset-level lexicons can be used for sentiment analysis at the broader levels of sentences and documents (Hung and Chen, 2016; Pamungkas and Putri, 2017). These works find that using synset lexicons in conjunction with word sense disambiguation techniques for English texts results in more precise sentiment predictions than those achieved using word-level lexicons. Similar improvements were observed using synset lexicons to classify non-English text as well (Denecke, 2008). Thus, the resource we develop can be used with these existing methods to perform downstream sentiment analysis tasks in multilingual settings.

3 Methodology

Our method to create a large set of sentiment-labeled synsets (SentiSynset) consists of two main stages. The first stage is to identify a set of emotional or polar synsets that we refer to as the *core*. In the second stage, this core is extended via WordNet relations that preserve sentiment. Our approach differs from prior works in that we create our core automatically, rather than manually. While assigning labels, we follow the precedent established in previously mentioned related works to map a synset to only one sentiment label.

3.1 Leveraging Word-Level Lexicons

To automatically develop the core of SentiSynset, we leverage existing multilingual sentiment lexicons created for sentiment analysis tasks at the *word* level. Sentiment labels for polysemous words may be inaccurate, due to different senses having different associated sentiments. We aim to resolve this ambiguity by leveraging translations, based on the observation that different senses of a word may translate differently. For example, *lick* translates into three distinct Dutch words, *ranselen*, *likken*, and *oplossen*, depending on the sense in which it is used. The sentiment labels associated with each Dutch translation can therefore be used to determine the appropriate label for each sense of *lick*.

While our method is bootstrapped from emotion lexicons, we make no assumptions about the languages or emotion inventories. Thus, our method is flexible and can be applied to other lexicons, potentially with larger vocabularies, or pertaining to specific domains. While emotional inventories vary, polarity labels are positive or negative.

Translating polysemous words into another language is not guaranteed to resolve *all* ambiguities that exist in word-level lexicons. For example, the two senses of *star* meaning "a celestial body of hot gases that radiates energy" and "someone who is dazzlingly skilled in a field" (definitions from WordNet) can both be translated as *estrella* in Spanish. This phenomenon is particularly prevalent among closely related languages; it is therefore advisable to perform translations into multiple languages with varying levels of similarity to English.

3.2 Developing the Core

Our method is designed to generate a core of high-precision synsets, which contain multiple words that are known to express a given sentiment. When labeling a synset, we consider the number of languages that contain sentimental lemmas belonging to the synset. For a lemma in a language other than English to be considered sentimental, it must share a sentiment label with an English lemma in the synset. For example, since the Indonesian lemmas in Figure 1 are associated with a disjoint set of emotions and polarity with respect to the English lemmas, they are disregarded when processing this synset.

To provide an emotional label (from a given emotion inventory) or polarity label (positive or negative), our method takes in a synset and finds all

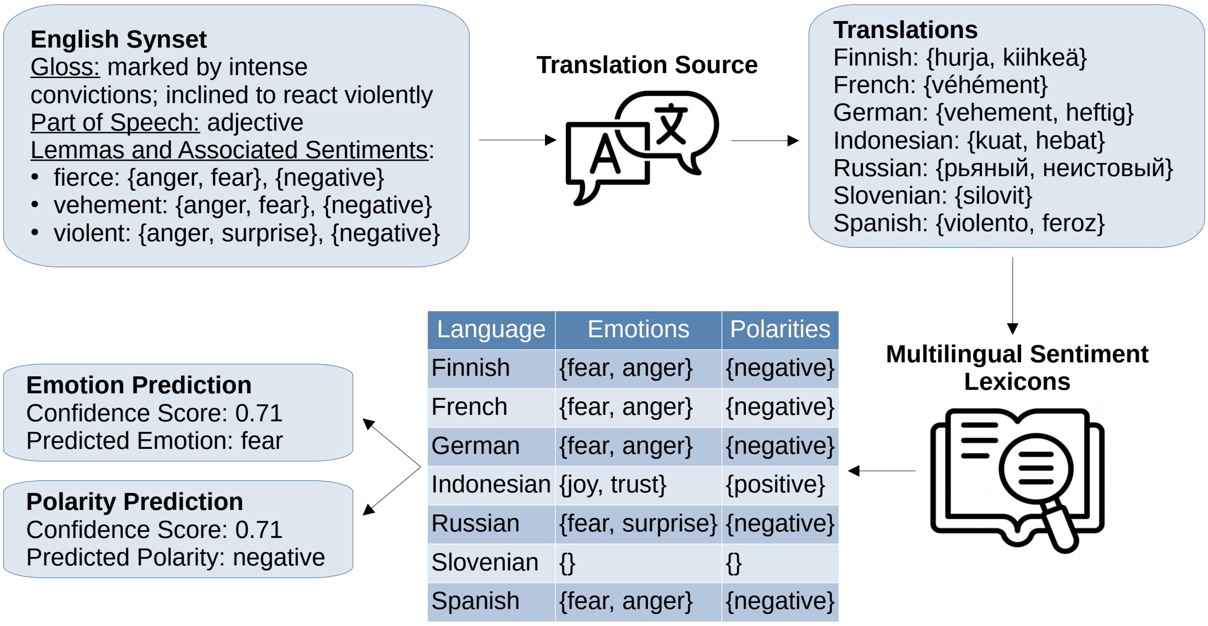

Figure 1: Illustration of performing emotion identification and polarity classification on a synset.

corresponding lemmas in the selected languages. We then determine which sentiments are associated with these lemmas using multilingual word-level lexicons. We finally associate the synset with the sentiment class which is associated with the highest number of translations. For example, since the synset in Figure 1 is associated with *fear* in 5 languages, *anger* in 4 languages, and *surprise* in 1 language, the synset is labeled with *fear*. Through a similar process, the synset is also associated with a negative polarity.

We calculate the confidence score of each candidate synset as the ratio of languages with sentimental lemmas to the total number of languages for which the synset has translations. For example, since the synset in Figure 1 has translations in seven languages, and lemmas that are considered emotional in five of the languages, it receives a confidence score of $5/7 \approx 0.71$.

Since each synset is assigned a single label, we proceed to break any ties that exist between sentiments that share the highest number of associated languages. For emotion identification, this is done by finding sentence embeddings for the gloss of the target synset and gloss for the most frequent sense of each of the top emotions. The synset is identified with the single emotion that has the most similar sentence embedding. For polarity identification, when a synset is associated with positive and negative polarities in the same number of languages, a similar process using sentence embeddings is applied to break the tie. We perform a comparison to this embedding-based approach as a baseline in Section 5.2.

3.3 Extending the Core

To expand the set of core synsets, we leverage WordNet's graph-based structure, which connects synsets through both semantic and lexical relations. Specifically, we propagate sentiment labels from the core to neighboring synsets via sentiment-preserving relations. If a synset is related to multiple core synsets with differing sentiments, we resolve this conflict with the embedding-based tie-breaking algorithm described in Section 3.2. In order to maintain high precision, we do not apply this procedure recursively or transitively.

We adopt the comprehensive set of sentiment-preserving relations used by WordNet-Affect, which differs slightly from the one used by SentiSense, and contains the following WordNet relations: *antonym*, *similar to*, *derived from*, *pertains to*, *attribute*, and *also-sees*. For example, the synset in Figure 1 is classified as having negative polarity, and is associated with the emotion of fear. The synset containing the adverbial sense of *fiercely* is related to this synset by the WordNet *pertains to* relation, and so is also labeled with negative polarity and the emotion of fear.

The relation of *antonymy* is unique in that it connects synsets that convey the *opposite* rather than identical sentiments. We follow Plutchik (1962) by identifying the following pairs of antonymic emotions: *anger/fear, anticipation/surprise, disgust/trust, and joy/sadness.* If a core synset is labeled with one of these sentiments, its antonyms are labeled with the opposite sentiment. Similarly, if a synset is labeled with positive or negative polarity, its antonyms are labeled with the opposite.

4 Experimental Setup

In this section, we provide details of our implementation and the datasets that we use.

4.1 Datasets

The NRC Word-Emotion Association Lexicon (EmoLex) (Mohammad and Turney, 2010, 2013), is a word-level sentiment lexicon which contains 14,182 English words tagged with emotional and polar labels by human annotators. Of those words, 4,454 are tagged with one or more of Plutchik's 8 fundamental emotions: *anger, anticipation, disgust, fear, joy, sadness, surprise*, and *trust.* As well, 5,543 of these words are tagged with positive and/or negative polarity. EmoLex was originally developed in English but has since been translated into 108 different languages. It is these translations that we use as our multilingual sentiment lexicons.

To evaluate the quality of SentiSynset, we construct both a validation set and a test set, each containing 1,000 synsets. Each set includes a random sample of 500 synsets from the SentiSense resource; these constitute the sentimental instances. Each also includes a random sample of 500 synsets that have no emotional or polar lemmas according to EmoLex or the LIWC dictionary (Pennebaker et al., 2001); these provide non-sentimental instances. We ensure that the validation and test sets are disjoint.

4.2 Synsets and Translations

The core of SentiSynset is found by applying the multilingual method described in Section 3.2 to all 117,659 WordNet synsets for the two independent tasks. We use the NLP library spaCy[2] to obtain sentence embeddings (c.f., Section 3.2).

Our method also requires a way of obtaining, for each synset, a set of words in various languages which lexicalize the concept to which that synset corresponds; for brevity, we refer to these multilingual terms as translations. We use translations for WordNet synsets in 20 languages covered by EmoLex: Chinese, Dutch, Estonian, Finnish, French, German, Greek, Indonesian, Korean, Lithuanian, Norwegian, Polish, Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, Turkish, and Ukrainian.

During development, we considered two translation sources. The first set of translations comes from the multilingual lexical database BabelNet (Navigli and Ponzetto, 2010). BabelNet was built by integrating various large lexical databases such as WordNet, Wikipedia, and Open Multilingual WordNet among others, alongside machine translation. We make use of BabelNet version 5.1, which covers over 500 languages; however, it does not contain translations for every synset in every language. On average, each of the selected 20 languages has BabelNet translations for 70.7% of all WordNet synsets. WordNet synsets have BabelNet translations in 14 of the selected languages on average, and 99.9% of all WordNet synsets have a BabelNet translation in at least one of the 20 selected languages.

The second set of translations comes from Google Translate (GT). To obtain sense-accurate translations, we translate an example sentence associated with the synset. WordNet provides such sentences for some synsets. For synsets without examples, we construct an example using the WordNet gloss. For instance, for the synset in Figure 1, we would construct the following sentence: "to be *fierce* is to be marked by intense convictions; inclined to react violently." Note that the templates used to construct sentences differ slightly depending on the synset's part of speech. We compile all the English sentences together and use the document translator on the GT online interface to attain the sentence translations. We then use the alignment system SimAlign (Jalili Sabet et al., 2020) to find the translation of the target word in the translated sentences. GT provides translations for every synset in every language, but not all translations are correct.

After obtaining translations from both BabelNet and GT, we lemmatize the translations using the Simplemma[3] library This step is skipped for languages such as Chinese and Korean where lemmatization is not applicable.

We consider four approaches to obtaining synset translations: GT alone, BabelNet alone, BabelNet supplemented with GT (i.e., BabelNet is used if

[2]https://spacy.io

[3]https://github.com/adbar/simplemma

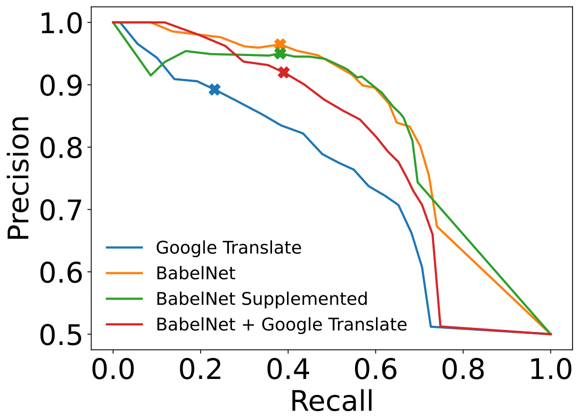

Figure 2: Precision-recall curve across various translation sources. Each curve is marked with a point corresponding to a high-confidence threshold of 0.70.

Sentiment	#Synsets
Anger	1891
Anticipation	1192
Disgust	1078
Fear	1939
Joy	1048
Sadness	1877
Surprise	391
Trust	3013
Positive	7081
Negative	8486

Table 1: Number of synsets associated with different sentiments in SentiSynset.

translations exist for a synset, otherwise GT is used), and the union of BabelNet and GT. On the validation set, we calculate the 11-point interpolated average precision (11-PIAP) (Manning et al., 2008) for each of these approaches and find that BabelNet alone results in the highest 11-PIAP of 83.9%, while GT, BabelNet Supplemented, and the union of BabelNet and GT result in the 11-PIAP's of 74.5%, 82.8% and 70.5% respectively. We also considered the precision and recall scores that the four translation approaches achieve on the validation set in the task of detecting emotional synsets (see Figure 2). Therefore, we use the best-performing method of BabelNet alone as the source of synset translations.

5 Results

In this section, we evaluate our method's performance in the desired tasks and discuss the newly created resource.

5.1 Core Synsets

To determine the confidence threshold of the method, we look at the experimental results of using BabelNet translations on the validation set (as shown in Figure 2). Since we want high-precision predictions, we choose the confidence threshold with a precision above 0.95 which has the highest recall. We find that a confidence threshold of 0.70 satisfies this condition, and thus all synsets that are predicted to be sentimental with a confidence score of 0.70 or above are added to the core of SentiSynset.

With the high-confidence threshold set, running the method on all WordNet synsets for the tasks of emotion identification and polarity classification results in a core containing 6,056 synsets that are predicted to be emotional and a core containing 8,519 synsets that are predicted to be polar. After extending these cores through the use of sentiment-preserving WordNet relations, SentiSynset contains a total of 12,429 emotional synsets and a total of 15,567 polar synsets. Information regarding the distribution of sentiments and parts of speech in these synset sets is shown in Tables 1 and 2.

5.2 Emotion Identification

To evaluate the quality of our newly constructed emotion resource, we measure the proportion of correct sentiment labels. We consider synsets in the intersection of our emotion resource and the test set. If a synset labeled as sentimental is in the intersection, we consider this a true positive. If a non-sentimental synset is in the intersection, we consider this a false positive. Using these classifications, we find that our method achieves a precision of 96.0% and a recall of 57.2% in the task of detecting emotional synsets.

To determine the accuracy of the emotion labels given by our method, three native English-speaking authors of this paper independently annotated all true positive synsets in the test set with one of the 8 fundamental emotions. The annotators achieved an average pairwise Cohen's kappa coefficient of 0.60, suggesting substantial agreement between the annotations. Similarly, at least 2 annotators agreed on a label for 92.3% of the synsets, and all 3 annotators agreed on a label for 53.3% of the synsets. For the 7.7% of synsets that all three annotators disagreed

POS	Emotional	Polar
Adjective	4531	5879
Adverb	144	237
Noun	5301	6464
Verb	2453	2987
Total	12,429	15,567

Table 2: Number of synsets associated with different parts of speech (POS) in SentiSynset.

Method	Emotion	Polarity
Random EmoLex	34.5	82.0
Sentence Embeddings	41.5	85.4
SentiWordNet	–	91.3
ChatGPT	79.0	93.1
Ours	**82.4**	**95.8**

Table 3: Accuracy of our method versus other approaches on the test set (in %).

on, the annotators were asked to reconsider their labels after being shown the emotions assigned to the synsets by the other annotators. Once all synsets had a single emotion that the majority of annotators agreed upon, these emotions were taken as the true labels.

We compare our method to several approaches. As a baseline, we find all emotions related to the English lemmas of a synset in EmoLex, then label a synset with a random emotion from this set. *Sentence Embeddings* takes these same emotions found in the English lexicon, computes sentence embeddings for the gloss of the target synset and for the gloss of the most frequent sense of each of these emotions, and labels the synset with the most similar emotion. We also prompt GPT-3.5 (Brown et al., 2020) to provide emotional labels for the synsets based on gloss and the lemmas. Finally, we classify synsets with the emotion labels assigned by our multilingual method. The accuracy of these different approaches can be found in Table 3, and we find that our method achieves the best performance.

No comparisons are made between the emotion labels assigned by our method and those of an existing resource because of the incongruent emotional inventories used between different synset-level resources; neither WordNet-Affect nor SentiSense uses Plutchik's 8 fundamental emotions as we do.

5.3 Polarity Classification

We evaluate the quality of our newly constructed polarity resource through a similar process used to evaluate our performance in emotion identification. When comparing the intersection of the test set and our polarity resource, we find that our method achieves a precision of 92.0% and a recall of 67.0% in the task of detecting polar synsets. We compare our polarity resource to SentiWordNet. Since SentiWordNet assigns synsets polarity scores in the range [0.0, 1.0], we assign synsets a single polarity label based on these scores. We do so by associating a synset with the polarity category (positive, negative, or objective) with the highest score. For the intersection between the test set and polar SentiWordNet synsets, SentiWordNet achieves a precision of 91.6% and recall of 41.6%.

To determine the accuracy of the polarity labels given by different methods, we convert the emotional labels given to the synsets by SentiSense to polarity labels. The emotions of *calmness, hope, joy, like,* and *love* are associated with positive polarity, while *anger, despair, disgust, fear, hate,* and *sadness* are associated with negative polarity. We disregard synsets associated with the emotions of *ambiguity, anticipation,* or *surprise* since synsets labeled with these emotions are not strongly correlated to either polarity. These emotion-to-polarity mappings, alongside equivalent polarity labels, are considered the true positives.

The methods that we compare for polarity classification are similar to those for emotion identification. Our baseline is to assign synsets with a random polarity that is associated with the English lemmas in EmoLex. We also compare to Sentence Embeddings, SentiWordNet, and ChatGPT. As shown in the rightmost column of Table 3, our method again achieves the best performance.

5.4 Polysemous Words

We investigate how well our method can resolve the ambiguity of polysemous words. To do so, we identify pairs of synsets in the test set that share a lemma but have opposite sentiments (polar and non-polar, emotional and non-emotional). Since our method focuses on precision over accuracy, we only consider pairs of synsets that share a polysemous word when at least one of the synsets is predicted to be sentimental.

We find 18 pairs of synsets with polar and non-polar labels in the test set, and our method provides correct classifications for both senses with 94.4% accuracy. The only pair of synsets that the method fails to correctly classify contains the polar and non-polar senses of *sublime* meaning "of high moral or intellectual value" and to "vaporize and then condense right back again" (WordNet). Our method identifies both senses as being positive, while this is only true for the first sense.

Our method is 100.0% accurate on 10 pairs of synsets with emotional and non-emotional labels that exist in the test set. For example, given the senses of *plume* meaning to "be proud of" and "(of a bird) to clean with one's beak" our method correctly identifies the first one as emotional and associated with joy, and the second one as non-emotional.

6 Error Analysis

In this section, we investigate incorrect labels produced by our method and discuss possible causes and solutions for such errors.

6.1 Parallel Polysemy

Our method struggles to correctly label concepts that exhibit parallel polysemy across many of the selected languages. For example, two nominal senses of *resistance* meaning "the action of opposing something that you disapprove or disagree with" and "a material's opposition to the flow of electric current; measured in ohms" share the same translation in French (*résistance*), German (*widerstehen*), Polish (*opór*), and 12 other languages. This causes the first sentimental sense to be viewed the same as the second non-sentimental sense, leading to an incorrect classification.

Although our selected languages do not all come from the same language family, the majority of them are European. This relatedness means they are more susceptible to exhibiting parallel polysemy than if we were to use more non-European languages. However, most non-European languages have considerably fewer lexical resources available than European languages, even for widely spoken non-European languages. For example, Estonian has 1.1 million speakers while Yoruba has 44.0 million speakers (Eberhard et al., 2023); nevertheless, BabelNet has translations available in Estonian for 6.4 times as many synsets than those that are available in Yoruba.

If synset translations were more readily available for languages such as Yoruba or Igbo, parallel polysemy would present less of a problem. Regarding the example of *resistance* above, the two senses would be translated to *atako* and *resistance* in Yoruba. In Igbo, the first sense translates to *iguzogide* while the second does not translate to a single word. Thus, translations from either language would help disambiguate the sentiment of the senses.

6.2 EmoLex Errors

The multilingual versions of EmoLex are translations of the original English EmoLex, so some translation errors exist in these translated lexicons. Words are typically translated as their most frequent sense (MFS), which causes issues when the MFS is non-sentimental. For example, the MFS of *waffle* is the non-sentimental nominal sense meaning "pancake batter baked in a waffle iron." However, *waffle* is considered sentimental in English Emolex due to the verbal sense meaning to "pause or hold back in uncertainty or unwillingness" (WordNet). When EmoLex is translated to other languages, *waffle* is translated as the non-sentimental MFS, but retains the sentiments associated with the verbal sense. Therefore, errors arise such as the Slovak word *vafle* being associated with the emotion of sadness and a negative polarity, despite the word referring strictly to the food item. These translation errors in EmoLex result in the MFS of *waffle* being incorrectly classified as sentimental.

6.3 Subjectivity

Other errors arise from the inherently subjective nature of the given tasks. It is very possible that one person may view a synset as sentimental, while another person views the same synset as non-sentimental. This causes issues when the method correctly determines which word sense EmoLex references, but the accuracy of the EmoLex annotation itself is debatable. For example, *bee* is associated with the emotions of anger and fear in EmoLex, with this annotation most likely referring to the MFS of the word meaning "any of numerous hairy-bodied insects including social and solitary species" (WordNet). Since the method bases its classifications on EmoLex, this sense of *bee* is associated with fear. However, some people may feel that this classification is inappropriate, instead viewing the synset as non-emotional. This opposing view is supported by the fact that *wasp* is not

Language Pair		Emotion	Polarity
Igbo	Yoruba	0.446	0.360
Chinese	Igbo	0.410	0.166
Chinese	Yoruba	0.390	0.401
Polish	Chinese	0.334	0.105
Polish	Igbo	0.353	0.354
Polish	Yoruba	0.292	0.355

Table 4: Cohen's kappa coefficient between emotion and polarity labels for different languages.

associated with any emotions in EmoLex, despite this term being very similar to *bee*.

Subjectivity is also influenced by cultural differences. While an English-speaking annotator labeled *bee* with the negative emotions of anger and fear in EmoLex, people from other cultures may associate bees with positive emotions as they are often considered hard-working creatures. This contrasting sentiment of the word that exists in English may be projected onto sentiment lexicons in other languages because of the virtual hegemony of English resources.

6.4 Cultural Differences

Our multilingual method hinges upon the idea that the sentiments associated with synsets tend to be universal across languages and cultures. However, the *bee* example demonstrates that this is not always the case. We therefore perform a multilingual analysis to quantify the influence of cultural differences on synset classifications.

We utilize plWordNet (Maziarz et al., 2016), a Polish wordnet that contains over 30,000 word senses that have been manually annotated with emotion and polarity labels (Zaśko-Zielińska et al., 2015). Of these labeled synsets, many have mappings onto Princeton WordNet, thus allowing us to investigate the effect of cultural differences on synset labels. There are 1,729 polar synsets and 1,506 emotional synsets that have sentiment labels in both our resource and plWordNet. The polarity and emotional labels have 94.9% and 73.3% agreement, respectively, between the two resources.

Authors of this paper who are native Chinese, Igbo, Polish, and Yoruba speakers labeled 40 polar synsets and 60 emotional synsets, which are among those that plWordNet and SentiSynset disagree on. The annotators were provided with the lemmas and glosses of synsets in their native language, with this information coming from BabelNet when available and Google Translate when not. For the Polish annotator, lemmas and glosses for all synsets were available from plWordNet.

The results of this experiment are shown in Table 4. The average Cohen's Kappa coefficient between the annotations for Polish and the three other languages (the last three rows of the table) are 0.326 and 0.271 for emotion and polarity, respectively. The Cohen's Kappa coefficient between our Polish annotator and plWordNet are 0.387 and 0.203 for emotion and polarity, respectively. Thus, the agreement between our Polish annotator and plWordNet for these contentious synsets is at a similarly low level as the agreement between annotators from different cultures.

7 Conclusion

We have presented a novel method that leverages multilingual translations to shift the sentimental classifications of word-level lexicons from words to synsets. The method is sufficiently general to be applied to the related yet independent tasks of emotion identification and polarity classification. The method outperforms existing methods used to automatically construct resources for the task of polarity classification. With our method, we constructed SentiSynset, which is substantially larger than comparable English sentiment resources. The large number of labeled synsets, and the high precision of labeling demonstrate the method's usefulness. The new resource can be paired with word-sense disambiguation techniques for the downstream task of sentiment analysis at the level of sentences or documents. Since our method is not dependent on EmoLex, it could also leverage information from multiple word-level lexicons, which could further improve the quality and size of SentiSynset.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Alberta Machine Intelligence Institute (Amii).

References

Muhammad Abdul-Mageed and Lyle Ungar. 2017. EmoNet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 718–728, Vancouver, Canada. Association for Computational Linguistics.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *CoRR*, abs/2005.14165.

Jorge Carrillo-de Albornoz and Laura Plaza. 2013. An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *Journal of the American Society for Information Science and Technology*, 64(8):1618–1633.

Yanqing Chen and Steven Skiena. 2014. Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 383–389, Baltimore, Maryland. Association for Computational Linguistics.

Fermín Cruz, José Troyano, Beatriz Pontes, and F. Javier Ortega. 2014. Ml-senticon: A multilingual, lemma-level sentiment lexicon. *Procesamiento de Lenguaje Natural*, 53:113–120.

Jorge Carrillo de Albornoz, Laura Plaza, and Pablo Gervás. 2012. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3562–3567, Istanbul, Turkey. European Language Resources Association (ELRA).

Kerstin Denecke. 2008. Using sentiwordnet for multilingual sentiment analysis. In *2008 IEEE 24th International Conference on Data Engineering Workshop*, pages 507–512.

Luca Dini and André Bittar. 2016. Emotion analysis on Twitter: The hidden challenge. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3953–3958, Portorož, Slovenia. European Language Resources Association (ELRA).

David M. Eberhard, Gary F. Simons, and Charles D. Fennig, editors. 2023. *Ethnologue: Languages of the World*, twenty-sixth edition. SIL International, Dallas, Texas.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, page 168–177, New York, NY, USA. Association for Computing Machinery.

Chihli Hung and Shiuan-Jeng Chen. 2016. Word sense disambiguation based sentiment lexicons for sentiment classification. *Knowledge-Based Systems*, 110:224–232.

Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. SimAlign: High quality word alignments without parallel training data using static and contextualized embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1627–1643, Online. Association for Computational Linguistics.

Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing*, 2nd edition. Prentice Hall.

Tuomo Kakkonen and Gordana Galić Kakkonen. 2011. SentiProfiler: Creating comparable visual profiles of sentimental content in texts. In *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, pages 62–69, Hissar, Bulgaria. Association for Computational Linguistics.

Jasy Suet Yan Liew and Howard R. Turtle. 2016. Exploring fine-grained emotion detection in tweets. In *Proceedings of the NAACL Student Research Workshop*, pages 73–80, San Diego, California. Association for Computational Linguistics.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. 2008. *Introduction to information retrieval*. Cambridge University Press.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. plWordNet 3.0 – a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268, Osaka, Japan. The COLING 2016 Organizing Committee.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Saif Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation* *of Emotion in Text*, pages 26–34, Los Angeles, CA. Association for Computational Linguistics.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.

Endang Wahyu Pamungkas and Divi Galih Prasetyo Putri. 2017. Word sense disambiguation for lexicon-based sentiment analysis. In *Proceedings of the 9th International Conference on Machine Learning and Computing*, ICMLC '17, page 442–446, New York, NY, USA. Association for Computing Machinery.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 271–278, Barcelona, Spain.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Robert Plutchik. 1962. *The emotions: Facts, theories, and a new model*. Random House.

Tiberiu Sosea and Cornelia Caragea. 2020. CancerEmo: A dataset for fine-grained emotion detection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8892–8904, Online. Association for Computational Linguistics.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet affect: an affective extension of WordNet. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Carlo Strapparava, Alessandro Valitutti, and Oliviero Stock. 2006. The affective weight of lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Peter Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Monika Zaśko-Zielińska, Maciej Piasecki, and Stan Szpakowicz. 2015. A large Wordnet-based sentiment lexicon for Polish. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 721–730, Hissar, Bulgaria.