

# Homonymy and Polysemy Detection with Multilingual Information

Amir Ahmad Habibi, Bradley Hauer, Grzegorz Kondrak

Alberta Machine Intelligence Institute, Department of Computing Science

University of Alberta, Edmonton, Canada

{amirahmad, bmhauer, gkondrak}@ualberta.ca

## Abstract

Deciding whether a semantically ambiguous word is homonymous or polysemous is equivalent to establishing whether it has any pair of senses that are semantically unrelated. We present novel methods for this task that leverage information from multilingual lexical resources. We formally prove the theoretical properties that provide the foundation for our methods. In particular, we show how the *One Homonym Per Translation* hypothesis of Hauer and Kondrak (2020a) follows from the synset properties formulated by Hauer and Kondrak (2020b). Experimental evaluation shows that our approach sets a new state of the art for homonymy detection.

## 1 Introduction

A word with multiple senses considered to be semantically ambiguous. In WordNet (Fellbaum, 1998), a word is semantically ambiguous if and only if it occurs in more than one synset. There are two types of word ambiguity: Two senses of a word are in the relation of *polysemy* if they are semantically related. For example, the WordNet senses `bank#n#2` “financial institution” and `bank#n#9` “financial building” are semantically related. Two senses of a word which are not semantically related are in the relation of *homonymy*. Continuing the example, `bank#n#1` “sloping land” is not related to `bank#n#2` “financial institution.” A word is *homonymous* if and only if it has a pair of senses in the homonymy relation. A word which is ambiguous, but not homonymous, is *polysemous*. *Polysemy classification* (Utt and Padó, 2011), or *homonymy detection*, is the task of automatically deciding whether a given ambiguous word is homonymous or polysemous. In this paper, we develop and present novel methods for this task.

Homonymy detection is vital to the tasks of defining sense inventories and clustering fine-grained senses (Navigli, 2006; Hovy et al., 2006). Distinguishing between homonymous and polysemous words is a core problem in lexicography (Mel’čuk, 2013). It has also been a subject of study in psycholinguistics (Brown, 2008). Consistent with the well-known tendency for distinct senses of a word to translate differently in other languages (Resnik and Yarowsky, 1999), Liu et al. (2018) show that special processing of homonymous words can improve neural machine translation.

Deciding whether a given word is homonymous or polysemous typically requires extensive manual effort and consultation of hand-crafted resources, as the intuitions of native speakers alone are not sufficient. WordNet does not contain homonymy information. Liu et al. (2018) rely upon a list of homonymous words obtained from a Wikipedia article, which is not reliable. Rice et al. (2019) manually identify a set of homonyms using a dictionary. Noting the lack of any existing homonymy resources of sufficient quality and coverage, Hauer and Kondrak (2020a) manually construct a list of homonyms from English etymological dictionaries, and map these homonyms to WordNet senses. While the list is not exhaustive, it provides a benchmark to evaluate homonymy detection methods.

We adopt a graph-based approach, constructing a sense graph for an input word using WordNet, and a multilingual extension, BabelNet (Navigli and Ponzetto, 2012). In such a graph, vertices are senses, while edges represent semantic relatedness. We make the simplifying assumption that the polysemy relation is transitive. Thus, any pair of senses which are connected in the sense graph are semantically related. Furthermore, if the graph has more than one connected component, the word is homonymous, as it has at least one pair of unrelated senses. Thus, the task of identifying homonymous

words is reduced to task of deciding the pairwise semantic relatedness of senses. We present a variety of methods for this sub-task, which leverage both monolingual and multilingual information. We formally prove the theoretical properties that provide the foundation for our methods. In particular, we show how the *One Homonym Per Translation* (OHPT) hypothesis of Hauer and Kondrak (2020a) follows from the synset properties formulated by Hauer and Kondrak (2020b).

The results of our experiments set a new state of the art for the task of homonym detection, outperforming the prior work of van den Beukel and Aroyo (2018). On a balanced dataset of homonymous and polysemous words, we achieve a 12% improvement in  $F_1$  score. We also investigate which combination of translation languages yields the best overall results.

## 2 Related Work

Homonymy detection has been investigated in natural language processing. Utt and Padó (2011) propose a statistical model which computes a polysemy index on the scale between homonymy and polysemy. Their work is not comparable to ours, as it depends upon an additional sources of ontological information. More recently, van den Beukel and Aroyo (2018) detect homonymous words for the task of humor recognition. Their method uses WordNet path similarity (Pedersen et al., 2005) and the textual similarity between synset definitions. Their definition of homonyms is broader than ours, including distinct word forms with identical pronunciations (homophones).

Homonymy detection has also been studied in psycholinguistics. Beekhuizen et al. (2018) distinguish between monosemous, polysemous, and homonymous words using word embeddings and contextual embeddings. The idea is that the embedding of a monosemous word should be closer to the embedding of its context than a polysemous word, which should in turn exhibit greater similarity to its context compared to a homonymous word. Rice et al. (2019) extract a list of 534 homonymous words from the Wordsmyth dictionary, and annotate them manually in sentential contexts. These resources are then used to analyze the relative frequencies of these homonyms.

As homonymous words are exactly those with semantically unrelated senses, the study of homonymy is closely related to the study of se-

semantic relatedness between word senses. To this end, Dyvik (2004) presents methods aimed at automatic construction of a WordNet-like resource using information extracted from parallel text corpora. This involves the use of translation information to induce semantic fields, which partition senses according to their semantic relatedness. Van der Plas and Tiedemann (2006) identify semantic relations between words using distributional information extracted from corpora, and show that leveraging multilingual data yields substantial improvements for the task of detecting synonymous words.

## 3 Background

Our work builds upon recent investigations into the linguistic phenomena of sense, homonymy, polysemy, synonymy, and translation. Therefore, we begin with a review of the relevant terminology, definitions, and general background knowledge.

### 3.1 Homonyms and Homonymous Words

Hauer and Kondrak (2020a) provide definitions of terms relevant to homonym detection. A *lexeme* is a single entry of a word in a lexicon (Jurafsky and Martin, 2008). A *word* is a basic written form which represents one or more lexemes. Each lexeme has at least one *sense*, corresponding to a use of the associated word to express a single lexicalized concept. The senses of a lexeme all relate to a single general meaning, and therefore they are all semantically related and so polysemous (Murphy and Koskela, 2010). Contrariwise, senses of a single word, but of distinct lexemes, are unrelated; the fact that a single word represents both lexemes may be entirely coincidental. Unrelated senses of a single word are *homonymous*; a word with a pair of homonymous senses is likewise called *homonymous*, and the lexemes it represents are called *homonyms*. Equivalently, a word is homonymous if it represents more than one lexeme. When it is clear that we are referring to a word, we can refer to a homonymous word simply as a homonym.

A classic example of a homonym is *bank*. This word represents two lexemes, referring generally to a repository (as in “bank account”), or to a slope (as in “river bank”). Each lexeme has multiple senses; for example the “slope” lexeme has senses expressing the concept of a shore, and that of an aircraft maneuver. A word is polysemous if it has multiple senses, but only one lexeme, i.e. if all of its senses are semantically related.

*Parallel homonymy* exists when two different words lexicalize the same pair of unrelated concepts. Parallel homonymy may exist between words in the same language, or in different languages. For example, both the English words *set up*<sup>1</sup> and *rig* have a pair of homonymous senses expressing the meanings “equip” and “manipulate”. A cross-lingual example involves the English *band* and Italian *banda*; each has a pair of homonymous senses expressing the meanings “ring” and “group”. Where cross-lingual parallel homonymy exists, two homonyms share at least one translation, which violates the OHPT hypothesis of Hauer and Kondrak (2020a).

### 3.2 Synsets, Wordnets, and Multi-Wordnets

Hauer and Kondrak (2020b) define synonymy as the relation of sameness of meaning. They note that it can be applied to various linguistic types – we can speak of synonymous words, synonymous senses, etc. – and that it can be conditional (*near-synonymy*) or absolute. The Princeton WordNet (Fellbaum, 1998) is composed of synonym sets, or *synsets*. Following prior work, we use the common noun *wordnet* to refer to any resource structured analogously to Princeton WordNet. A synset is a set of words which are all pairwise near-synonyms; each synset corresponds to a lexicalized concept, which each word in the synset can be used to express. There exists a one-to-one correspondence between the senses of a word, and the synsets which contain it; therefore, synsets induce a sense inventory. Synsets have the following properties (Hauer and Kondrak, 2020b):

1. *A word is monosemous iff it is in a single synset. A word is polysemous iff it is in multiple synsets.*
2. *Words are near-synonyms iff they share at least one synset. Words are absolute synonyms iff they share all their synsets.*
3. *Word senses are synonymous iff they are in the same synset.*
4. *Every word sense belongs to exactly one synset.*
5. *Every sense of a polysemous word belongs to a different synset.*

Multilingual wordnets (multi-wordnets) consist of multi-lingual synsets (multi-synsets). They are constructed either by adding words from other

languages to the (monolingual) synsets of a pre-existing wordnet, or linking synsets from multiple wordnets in different languages (Vossen, 1996). In any case, a multilingual synset can be viewed as set of words, each associated with a language, and capable of representing the lexicalized concept to which the multi-wordnet corresponds.

A wordnet facilitates the enumeration of the senses of a word, by identifying the concepts associated with the synsets containing the word. A multi-wordnet further enables the enumeration of the translations of a specific sense of a word, by retrieving the elements of the corresponding synset, excluding those from the same language as the word to be translated. Hauer and Kondrak (2020b) refer to this property as the *multi-wordnet assumption*: senses share a synset if and only if they are semantically equivalent.

## 4 Methods

In this section, we present our graph-based method for deciding whether a given word is homonymous or not. Under our definitions, this is equivalent to deciding whether the word has any senses that are semantically unrelated. Operating under the assumption that semantic relatedness of senses is symmetric and transitive, our strategy is as follows:

1. Enumerate the senses of the word.
2. If the word has only one sense, classify it as monosemous (and therefore not a homonym).
3. Identify pairs of senses that are semantically related.
4. Construct a graph whose vertices are senses of the input word, which are connected by an edge if they are semantically related.
5. Classify the word as homonymous if its graph has more than one connected component; otherwise, classify the word as polysemous.

In the semantic graph constructed by this procedure, adjacent senses are semantically related by definition, and connected senses are semantically related by transitivity. Therefore, the existence of more than one connected component implies that the word has semantically unrelated senses, and so is homonymous. Any sense inventory can be used to enumerate senses, and graph connectivity can be decided using a simple breadth-first search. All that remains is to establish methods of detecting whether two senses of a word are semantically related. We present five sufficient conditions for semantic relatedness between senses, one in each

---

<sup>1</sup>Following the example of WordNet, non-compositional multi-word units are considered words.

of the following five subsections. For each of these criteria, we describe the circumstances under which it detects semantic relatedness, and provide a theoretical argument for its soundness.

One strength of this method is its lack of dependence on any hyperparameters or additional training data. A vector-based method leveraging distributional semantics, for example, would necessarily depend on some continuous measure of semantic similarity. This in turn would require access to a large text corpus to learn distributional embeddings. In addition, a threshold value would need to be tuned, which would depend on the embeddings and the corpus. By avoiding such requirements, our method is easier to apply to arbitrary domains, and can be applied to any language which is represented in a multi-wordnet.

#### 4.1 Two Senses, One Translation (OHPT)

Our first method is an application of the OHPT hypothesis of Hauer and Kondrak (2020a). OHPT states that semantically unrelated senses of a word do not share any translations. It follows that if two senses of a word can be translated by a single word in another language, those senses are semantically related. We refer to this approach to detecting semantic relatedness as “two senses, one translation” or simply as OHPT.

The following theorem generalizes the OHPT hypothesis to account for the few exceptions found by Hauer and Kondrak (2020a):

**Theorem 1** (Two Senses, One Translation). *If two distinct senses  $x_1$  and  $x_2$  of a word  $x$  in one language can be translated by a word  $y$  in another language, then one and only one of the following condition holds:*

1.  $x_1$  and  $x_2$  are polysemous.
2.  $x$  and  $y$  exhibit parallel homonymy.

*Proof.* Distinct senses  $x_1$  and  $x_2$  belong to different multi-synsets (by synset property #5). Since they both can be translated by  $y$ , the two multi-synsets also contain senses  $y_1$  and  $y_2$  of  $y$ , respectively (by the multi-wordnet assumption).  $x_1$  and  $x_2$  are unrelated if and only if  $y_1$  and  $y_2$  are unrelated, as they express the same pair of concepts. Therefore, either all four senses are related, or  $x$  and  $y$  exhibit parallel homonymy.  $\square$

Theorem 1 is the principal theoretical result in this work, which provides a theoretical foundation

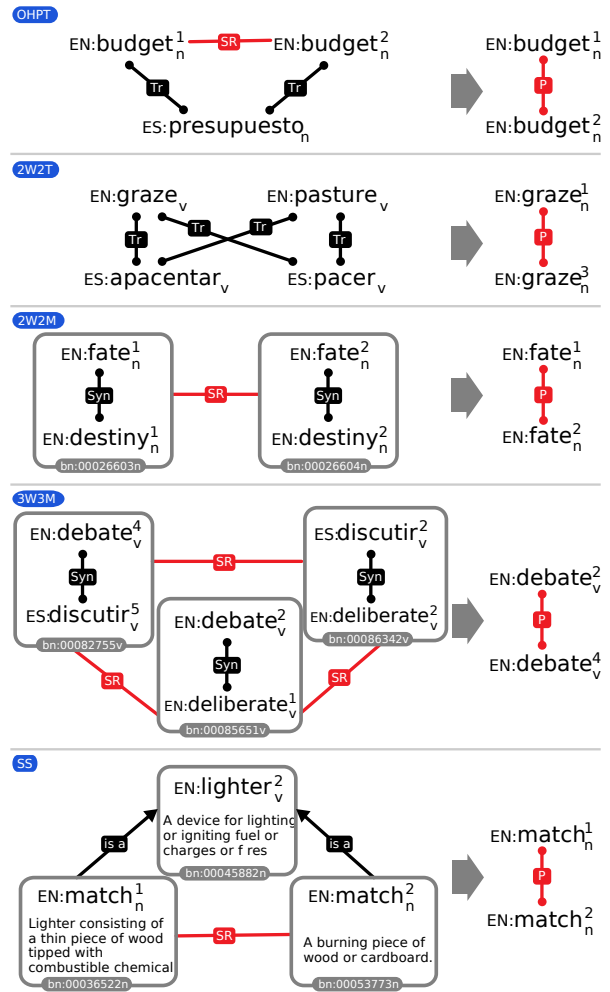


Figure 1: Examples that illustrate our methods. The semantic relations in black are provided by the resources. The polysemy relations in red are inferred by the methods.

for the OHPT hypothesis. In fact, all actual English-Italian exceptions to the hypothesis that Hauer and Kondrak (2020a) identify in their experiments involve parallel homonymy.

Operating under the assumption that parallel homonymy is rare<sup>2</sup>, we arrive at the criterion described above: senses which share a translation are related. The first part of Figure 1 shows an example application of OHPT.

#### 4.2 Two Words in Two Multi-synsets (2W2M)

Theorem 1 can be generalized to include pairs of words from the same language. This follows from the observation of Hauer and Kondrak (2020b)

<sup>2</sup>Hauer and Kondrak (2020a) find only three cases of parallel homonymy, across two languages of translation, using a database of 2759 homonyms.

that intra-lingual synonymy and cross-lingual synonymy are two views of the same phenomenon: semantic equivalence. As words in different languages share multi-synsets if they are mutual translations, so words in the same language share multi-synsets if they are near-synonymous. The following theorem captures this insight:

**Theorem 2** (Two Words in Two Multi-synsets). *If two distinct words  $x$  and  $y$  (either in the same or different languages) share two multi-synsets, then one and only one of the following condition holds:*

1. *The two pairs of senses of  $x$  and  $y$ , which correspond to those synsets, are polysemous.*
2.  *$x$  and  $y$  exhibit parallel homonymy.*

*Proof.* Senses that share a multi-synset are in the relation of absolute synonymy (Property 3). Thus, the two senses of  $x$  are unrelated if and only if the two senses of  $y$  are unrelated. Therefore, either all four senses are related, or  $x$  and  $y$  exhibit parallel homonymy.  $\square$

Interestingly, Theorem 2 provides a theoretical foundation for Heuristic #3 of Pericliev (2015) which states that the presence of distinct synonymous colexifiers in one language indicates polysemy rather than homonymy. Our theorem establishes that the only reason for the heuristic to fail is parallel homonymy. Since parallel homonymy is rare, we will assume that two senses of a given word are semantically related whenever their respective synsets or multi-synsets share another word. See Figure 1 for an example.

### 4.3 Three Words in Three Multi-synsets (3W3M)

Theorem 1, implies that, in the absence of parallel homonymy, if two senses of a word share a synset with a single word from another language, they are related. Theorem 2 removed the “from another language” clause. Theorem 3 extends this further, from two words sharing two multi-synsets to three words (in any combination of languages) sharing three multi-synsets.

**Theorem 3** (Three Words in Three Multi-synsets). *If three pairs of senses  $(x_1, y_1)$ ,  $(x_2, z_2)$ , and  $(y_3, z_3)$  of three different words  $x$ ,  $y$ , and  $z$  share three multi-synsets, respectively, then one and only one of the following conditions holds:*

1. *The three pairs of senses of  $x$ ,  $y$ , and  $z$  are polysemous.*

2. *At least two of the three words are homonymous.*

*Proof.* Without loss of generality, suppose that  $x_1$  and  $x_2$  are not semantically related, and that the words  $y$  and  $z$  are not homonymous. Then both  $y_1$  and  $y_3$ , and  $z_2$  and  $z_3$  are semantically related. By transitivity of semantic relatedness, this implies that  $x_1$  and  $x_2$  are also semantically related. Contradiction.  $\square$

This gives us our third criterion for establishing the semantic relatedness of a pair of senses of a word. As with the previous two theorems, given the rarity of homonymy, if the antecedent condition of the theorem is satisfied, the senses of each word involved are taken to be related. See Figure 1 for an example.

### 4.4 Two Words, Two Translations (2W2T)

Our next theorem is defined on words rather than senses.

**Theorem 4** (Two Words, Two Translations). *If two distinct words  $x$  and  $y$  in one language  $E$  both translate into two different words  $w$  and  $z$  in another language, then one and only one of the following conditions holds:*

1. *All senses involved in all those translation instances are semantically related.*
2. *At least two of the four words are homonymous.*

*Proof.* Without loss of generality, suppose senses  $x_1$  and  $w_1$  are mutual translations, and that neither is semantically related to any senses of  $y$  or  $z$ . Then, since  $x$  can also be translated as  $z$ , there exists a translation instance between another sense of  $x$ , call it  $x_2$ , and a sense of  $z$ , call it  $z_2$ . Since  $x_1$  and  $w_1$  are not semantically related to  $z_2$ ,  $x_1$  and  $x_2$  are not semantically related to each other (otherwise  $x_1$  and  $w_1$  would be related to  $z_2$  by transitivity through  $x_2$ ). So  $x$  is a homonymous word. A symmetrical reasoning leads to the conclusion that  $w_1$  and  $w_2$  are not semantically related to each other, so  $w$  is also a homonymous word.  $\square$

Given the rarity of homonymy (most ambiguous words are not homonymous), we again create a condition for semantic relatedness which assumes that the antecedent implies the first condition. Figure 1 shows an example application of this theorem.

## 4.5 Sibling Synsets (SS)

So far, the only semantic relations leveraged in our method are synonymy and translational equivalence. However, WordNet and comparable resources contain various other semantic relations between synsets, which could be used to infer semantic relatedness. In particular, we hypothesize that synsets which share a common hypernym or holonym are semantically related. We call this method Sibling Synsets (SS). Figure 1 illustrates an example of this method, with three synsets and their definitions.

## 4.6 UNION

We began this section by describing a method which classifies an ambiguous word as homonymous or polysemous by constructing a graph of its senses. In this graph, senses have an edge between them if and only if they are semantically related. This reduces the task of homonym detection to detecting pairwise semantic relatedness between senses of a single word. That is, given two senses of a word, are they senses of a single lexeme, or distinct lexemes?

We have presented five sufficient conditions for the semantic relatedness of senses: OHPT, 2W2M, 3W3M, 2W2T, and SS. Each of these criteria can be seen as adding a set of edges to the sense graph created by our method. It is, of course, possible to combine these methods by simply taking the union of the edge sets they return. This leads to a final criterion, **UNION**, which finds a pair of senses to be semantically related if any of the aforementioned five criteria do.

## 5 Experiments

Having provided theoretical support for our homonym detection methods in Section 4, we now empirically evaluate their ability to distinguish between homonymous and polysemous words. We start by describing our datasets and resources, as well as the method for obtaining translations of senses. We then present our results on both words and senses.

### 5.1 Data and Resources

To test our method, we create a balanced benchmark dataset for homonym detection consisting of 948 words, with 474 in each class. Note that the hand-crafted English resources described below are used for evaluation only, as our methods are

largely language-independent, being based on the BabelNet’s sense translation information.

We extract the positive instances, from the list of homonymous words released by Hauer and Kondrak (2020a). For each homonymous word, the list includes a partial mapping of its WordNet 3.0 senses to the individual homonyms. The original homonym-based sense clustering is both incomplete and noisy, due to the use of an automated pre-clustering procedure (Navigli, 2006) in the mapping process. We completed this mapping, ensuring that *all* senses of the included words are mapped to homonyms. We also manually corrected a number of errors in the mapping, where necessary. This new version of the resource has 474 homonymous words with a total of 1017 homonyms.<sup>3</sup> We make the corrected resource publicly available.

We also carefully select the negative (non-homonymous) instances. Strictly speaking, any word which has more than one sense, and which is not in any homonym list known to us, could be labeled a polysemous word and used as a negative example. However, to make the dataset more challenging, we take advantage of the manually-crafted and validated clustering of WordNet 3.0 senses released as part of the OntoNotes project (Hovy et al., 2006). We select 474 words at random from among the 3232 words which have more than one sense cluster in OntoNotes, but which are not among the homonyms described above. Thus, the dataset requires a homonym detection method to distinguish homonyms from words which not only have multiple senses in the fine-grained WordNet sense inventory, but also in the coarse-grained OntoNotes sense inventory.

We set aside 20% of these words (95 homonymous and 95 polysemous) as a test set, and use the rest for development. We compare our method to a simple baseline which simply predicts every word to be homonymous. We experimented with other baselines which threshold the number of WordNet senses of the input word, but they failed to consistently outperform this simple baseline.

### 5.2 Language Selection

Our criteria for semantic relatedness crucially depend upon sense translation information: words in other languages that can be used to express a given concept. We obtain this information from

<sup>3</sup>We exclude words whose homonyms have entirely disjoint parts of speech, e.g. *bear*, and words with only one homonym represented in WordNet 3.0, e.g. *wit*.

BabelNet multi-synsets. However, BabelNet is not a hand-crafted resource, and as such suffers from both coverage and accuracy problems. Specifically, many languages are only sparsely represented, and many automatically-generated translations are simply incorrect for a given sense. Since our method is based exclusively on the positive evidence for polysemy, considering more languages of translations improves the precision of homonymy detection, but decreases the recall, as many sense pairs are incorrectly identified as related. Thus, we attempt to identify a set of languages that yields the best trade-off in terms of  $F_1$  score on our development set.

Since it would be infeasible to test all possible combinations of hundreds of languages, we instead perform a heuristic search for a reasonably well-performing set of languages. We select our languages of translation from the 50 languages with the highest overall synset coverage in BabelNet. We then evaluate the performance of the OHPT method using each of these 50 languages. The resulting  $F_1$  scores of these development experiments provide a ranking of the languages, which we interpret as an estimate of its usefulness for our task. The top ten languages, in order, are Indonesian, Malay, Spanish, Catalan, Slovenian, Portuguese, Finnish, Italian, Romanian, and Croatian. It is difficult to interpret this ranking without language-specific knowledge, but we note that Indonesian and Malay are standardized varieties of an Asian *lingua franca*, while the others are European languages that share a substantial number of Greek and Latin roots.

We also experimented with combining translation information from multiple languages, whereby the evidence from any of them is accepted for establishing semantic relatedness of senses. In our development experiments, we found that the combination of Indonesian and Spanish yielded the best  $F_1$  score. Therefore, for our remaining experiments, when a criterion requires translation information, senses are considered semantically related if translation into Indonesian or Spanish provides evidence for semantic relatedness. We speculate that this combination is effective because they represent two very different languages that may complement each other.<sup>4</sup> It is likely that adding Malay or other European languages to the mix provides

<sup>4</sup>Resnik and Yarowsky (1999) note that distantly related languages seem to provide greater ability to resolve sense distinctions.

Method	Pre	Rec	$F_1$	Acc
Baseline	50.0	<b>100</b>	66.7	50.0
BA-2018	51.1	99.0	67.4	52.1
OHPT	74.5	83.2	<b>78.6</b>	<b>77.4</b>
UNION	<b>79.1</b>	71.6	75.1	76.3

Table 1: Homonym detection results on our test set, in terms of precision, recall,  $F_1$  score, and accuracy.

only a minimal gain in terms of diversity, at the cost of increasing the level of noise in the data.

### 5.3 Homonymy Detection

In our main evaluation experiment, we test two variants of our method, OHPT (Section 4.1) and UNION (Section 4.6), on our balanced test set of 190 words (95 homonymous, 95 polysemous). In addition to our naive baseline (Section 5.1), we compare our results against the method of van den Beukel and Aroyo (2018), which we denote as BA-2018.

The results are shown in Table 1. The baseline yields 100% recall and 50% precision and accuracy. Surprisingly, the BA-2018 method classifies almost all words as homonyms, which translates into only a small improvement in accuracy over the baseline. This attests to the difficulty of our dataset: highly-polysemous words with coarse-grained sense distinctions are not easy to distinguish from true homonyms.

Both OHPT and UNION easily outperform BA-2018. They identify far fewer false homonyms, resulting in a much higher precision. Interestingly, the OHPT criterion by itself gives better  $F_1$  score and accuracy than the UNION criterion. As UNION encompasses OHPT, the higher precision and lower recall of the latter are expected. This is because considering multiple criteria can only increase the connectivity of the sense graph. However, the higher  $F_1$  score of the simpler method is surprising. Based on these results, we conclude that the UNION criterion is best when precision is more important than recall, while the OHPT criterion is best when overall accuracy is desired.

### 5.4 Sense-Level Polysemy Detection

In our second experiment, we conduct a direct evaluation of the polysemy detection at the level of sense pairs. The task is deciding whether two senses of a single word are semantically related (positive classification) or unrelated (negative classification). This is different from our previous eval-

	Pre	Rec	F <sub>1</sub>	Acc
BA-2018	<b>92.2</b>	22.3	35.9	44.3
OHPT	86.5	65.8	74.7	68.8
UNION	81.8	<b>80.5</b>	<b>81.1</b>	<b>73.8</b>

Table 2: Sense-level SRC results on our test set, in terms of precision, recall, F<sub>1</sub> score, and accuracy.

uation on a word-level task of homonymy detection, where we used sense-level polysemy detection to create edges in the sense graph.

Table 2 presents the results on the same test set as in Section 5.3. Both OHPT and UNION substantially outperform the BA-2018 method in terms of F<sub>1</sub> and accuracy. This is consistent with our homonym detection results in Section 5.3: spurious positive classification at the sense level may lead to spurious *negative* classification on word-level homonym detection. Compared to OHPT and UNION, BA-2018 produces slightly fewer false positives, but many more false negatives. Consequently, our methods attain much higher recall for this task, which is in accordance with their much higher precision for homonym detection.

Unlike in the word-level experiment, UNION achieves better results than OHPT, because its much higher recall offsets its reduced precision. This establishes the utility of the various criteria developed in Section 4 for inclusion in UNION.

## 5.5 Error Analysis

The errors made by our homonymy detection methods can be divided into false positives and false negatives. While we used only Indonesian and Spanish translations in our English evaluation experiments, here we provide examples from languages with which we are more familiar.

False-positives, i.e. words incorrectly classified as homonymous, arise when two semantically related senses remain disconnected in the sense graph constructed by our method. We find that such cases are generally caused by data sparsity in BabelNet, which lacks many valid translations that could connect related senses in the graph. Another type of false positives is caused by *lexical gaps*, which occur when a language has no word or non-compositional phrase to express a given concept. For example, the sense of the polysemous adjective *seamless* glossed as “not having seams” corresponds to a lexical gap in Italian. Therefore, there is no translation in Italian that could relate

this sense to any other sense of *seamless*. An example of a lexical gap in English is the concept lexicalized by the Persian word پریروز (/pæri:ru:z/) “the day before yesterday.”

False-negatives are homonymous words for which our method finds evidence of relatedness between two unrelated senses. Such errors can be divided into three categories: spurious translation, incorrect sense-to-homonym mapping, and parallel homonymy. Below, we provide examples for each category.

First, many translations in BabelNet are incorrect because they were obtained from machine translation models. This spurious translation information, under criteria such as OHPT, can result in unrelated senses being classified as related. For example, the homonymous verb *shark* has two senses in BabelNet: “to act with trickery” and “to hunt sharks.” The French word *requin* shares both of the corresponding multi-synsets, which incorrectly implies that it translates both senses of *shark*. As a consequence, our method misclassifies these two senses as related.

Second, our homonym resource, even after manual cleaning, still contains some incorrect sense-to-homonym mappings, which are inherited from the automatic clustering of WordNet senses. For example, two unrelated senses of the noun *content*, (“the sum of what has been perceived” and “the state of being contented”) are incorrectly mapped to a single lexeme.

Finally, while parallel homonymy is rare, it does occur. As discussed in Section 4, parallel homonymy can create exceptions to our translation-based criteria for semantic relatedness, resulting in misclassifications. For example, two semantically unrelated senses of the English word *boil*, glossed as “the temperature of boiling” and “a painful sore,” can both be translated by the Persian word جوش /dʒu:ʃ/.

## 6 Conclusion

We have presented a novel approach to the problem of distinguishing between homonymy and polysemy. Our methods for establishing semantic relatedness leverage sense translation information from a multi-wordnet, and are supported by proofs constructed upon a formal theory of senses, synonymy, and translation. Our approach sets a new state of the art for the task of homonym detection. In the future, we would like to investigate stochas-



tic methods for this task, including random walks on semantic graphs, as well as the use of graph embeddings to compute the similarity of concepts in a dense vector space. To facilitate further research on homonymy detection, we make the augmented homonymy resource publicly available.<sup>5</sup>

## Acknowledgments

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

## References

- Barend Beekhuizen, Sasa Milic, Blair C Armstrong, and Suzanne Stevenson. 2018. What company do semantically ambiguous words keep? insights from distributional word vectors. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, pages 1347–1352.
- Susan Windisch Brown. 2008. Choosing sense distinctions for WSD: Psycholinguistic evidence. In *Proceedings of ACL-08: HLT, Short Papers*, pages 249–252, June.
- Helge Dyvik. 2004. Translations as semantic mirrors: from parallel corpus to wordnet. In *Advances in corpus linguistics*, pages 309–326. Brill Rodopi.
- Christiane Fellbaum. 1998. WordNet: An on-line lexical database and some of its applications. *MIT Press*.
- Bradley Hauer and Grzegorz Kondrak. 2020a. One homonym per translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7895–7902.
- Bradley Hauer and Grzegorz Kondrak. 2020b. Synonymy= translational equivalence. *arXiv preprint arXiv:2004.13886*.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60.
- Daniel Jurafsky and James H Martin. 2008. *Speech and Language Processing*. Prentice Hall.
- Frederick Liu, Han Lu, and Graham Neubig. 2018. Handling homographs in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1336–1345. Association for Computational Linguistics.
- Igor Mel’čuk. 2013. *Semantics: From meaning to text*, volume 2. John Benjamins.
- M. Lynne Murphy and Anu Koskela. 2010. *Key terms in semantics*. London: Continuum.
- Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.
- Roberto Navigli. 2006. Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 105–112, July.
- Ted Pedersen, Satanjeev Banerjee, and Siddharth Patwardhan. 2005. Maximizing semantic relatedness to perform word sense disambiguation. Technical report, Research Report UMSI 2005/25, University of Minnesota Supercomputing Institute.
- Vladimir Pericliev. 2015. On colexification among basic vocabulary. *Journal of Universal Language*, 63–93.
- Philip Resnik and David Yarowsky. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural language engineering*, 5(2):113–133.
- Caitlin A Rice, Barend Beekhuizen, Vladimir Dubrovsky, Suzanne Stevenson, and Blair C Armstrong. 2019. A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings. *Behavior research methods*, 51(3):1399–1425.
- Jason Utt and Sebastian Padó. 2011. Ontology-based distinction between polysemy and homonymy. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, pages 265–274.
- Sven van den Beukel and Lora Aroyo. 2018. Homonym detection for humor recognition in short text. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 286–291.
- Lonneke Van der Plas and Jörg Tiedemann. 2006. Finding synonyms using automatic word alignment and measures of distributional similarity. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 866–873. Association for Computational Linguistics.
- PJTM Vossen. 1996. Right or wrong: Combining lexical resources in the EuroWordNet project. In *M. Gellerstam, J. Jarborg, S. Malmgren, K. Noren,*

<sup>5</sup><https://webdocs.cs.ualberta.ca/~kondrak/homonyms.html>

*L. Rogstrom, CR Pappmehl, Proceedings of Euralex-96, Goetheborg, 1996, pages 715–728. Vrije Universiteit.*