One Sense per Translation

Bradley Hauer, Grzegorz Kondrak
Alberta Machine Intelligence Institute
Department of Computing Science
University of Alberta, Edmonton, Canada
{bmhauer,gkondrak}@ualberta.ca

Abstract

Word sense disambiguation (WSD) is the task of determining the sense of a word in context. Translations have been used in WSD as a source of knowledge, and even as a means of delimiting word senses. In this paper, we define three theoretical properties of the relationship between senses and translations, and argue that they constitute necessary conditions for using translations as sense inventories. The key property of One Sense per Translation (OSPT) provides a foundation for a translation-based WSD method. The results of an intrinsic evaluation experiment indicate that our method achieves a precision of approximately 93% compared to manual corpus annotations. Our extrinsic evaluation experiments demonstrate WSD improvements of up to 4.6% F1-score on difficult WSD datasets.

1 Introduction

Word sense disambiguation (WSD) is the task of classifying a word in context according to its sense. For example, given the context "the field was covered in green grass," a WSD system would need to classify *field* as having its "flat open land" sense, rather than its "area of study" sense. Throughout its history, WSD has been associated with translation (Weaver, 1949), as it is understood that different senses of a word may translate differently. For instance, in the above example, *field* could be translated into French as *champ*, but not as *domaine* (the latter could, however, translate the "area of study" sense of *field*). In this paper, we address the open question: *to what extent can a translation-based method improve modern WSD?*

This question is surely an important one: WSD remains an active area of research (Blevins and Zettlemoyer, 2020; Barba et al., 2021a,b), but despite the rapid improvements brought on by transformer-based (Vaswani et al., 2017) language models such as BERT (Devlin et al., 2019), substantial room for improvement remains (Maru et al., 2022). WSD has been used as a benchmark to compare and analyze transformer-based language models (Loureiro et al., 2021). It has also been shown to have applications to tasks such as translation (Liu et al., 2018), semantic parsing (Martínez Lorenzo et al., 2022), and metaphor detection (Maudslay and Teufel, 2022). New variants of the task are still being proposed, such as visual WSD, in which candidate senses are represented by images (Raganato et al., 2023). Clearly, the ability to map a word in context to an entry in a discrete lexical knowledge base remains relevant in natural language processing, for both human end users and downstream tasks.

Incorporation of translation information has been shown to be useful for both classic (Dagan et al., 1991) and modern (Luan et al., 2020) WSD methods. Despite such proof-of-concept works, current state-of-the-art WSD methods do not explicitly leverage translation, leaving a potential source of knowledge untapped. It is therefore of interest to the lexical semantics community to investigate the extent to which senses and translations correspond, and how this correspondence can be leveraged in practice.

Our investigation has the following structure: (1) We begin by clearly defining the theoretically "ideal" mapping between senses and translations. (2) We show that such mappings are rare in practice, even between unrelated languages, offering an explanation as to why translation-based WSD methods became less common as the field developed. (3) We posit that it is possible to improve supervised WSD performance by leveraging instances where the translation of a word *does* determine its sense. (4) We propose and evaluate a translation-based disambiguation method to test this hypothesis. (5) We discuss the relationship between various theoretical properties and synonymy and polysemy.

Our empirical results strongly support our hypothesis. A large-scale intrinsic evaluation of our

Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 442–454
November 1–4, 2023.

method using existing lexical knowledge bases shows that it achieves very high precision. Our extrinsic evaluation shows that synthetic training data produced by our method, when used to train a supervised model, can yield improvements in F1-score of up to 4.6% on difficult WSD benchmark datasets. We conclude that the explicit incorporation of contextual translations has great potential to improve WSD research, and lexical semantics research in general.

The principal goal of our paper is the examination of the sense-translation connection from both theoretical and empirical perspectives in a modern context. Thus our contributions are twofold: a theoretical analysis of the relationship between senses and translation, supported by empirical analysis; and a method for efficient, unsupervised, large-scale semantic annotation via translations, which yields substantial WSD improvements.

2 Related Work

The use of translations as a source of information about word senses rose to prominence in the 1990s, supported by the increasing availability of machine-readable multilingual resources. Brown et al. (1991) and Dagan et al. (1991) developed statistical approaches to WSD, with the former presenting a direct application to statistical machine translation. Gale et al. (1992) were the first to explicitly define WSD in terms of identifying the correct translation: they identify a set of six English words, each with two senses, with a one-to-one mapping between those senses and their French translations. This paradigm of translation-informed WSD influenced the landmark WSD works of Yarowsky (1995) and Schütze (1998), among others. By the late 1990s, translation was so prevalent in the WSD literature that Resnik and Yarowsky (1997) explicitly proposed "to restrict a word sense inventory to those distinctions that are typically lexicalized cross-linguistically."

Interest in translation in the WSD literature continued throughout the 2000s (Ide, 2000; Chan et al., 2007; Apidianaki, 2008), culminating in two SemEval-2010 shared tasks: cross-lingual lexical substitution (Mihalcea et al., 2010), and cross-lingual WSD (Lefever and Hoste, 2010). The former can be viewed as the task of finding translations for a word in a given context. In the latter, translations from word-aligned parallel corpora were used to create a "multilingual sense inventory". The dataset was limited to small lexical samples, and involved substantial manual-annotation effort for each tested language pair. Neither the exact annotation criteria nor the datasets themselves are available.

Yao et al. (2012) observed that prior work made conflicting assumptions about the correspondence between senses and translations. They consider the case where a single word e in a parallel corpus is aligned, in different contexts, with two different words, f_1 and f_2, in another language. They point out that some prior works, such as Lefever et al. (2011), assume that e is polysemous, with f_1 and f_2 translating distinct senses of e, while others, such as Bannard and Callison-Burch (2005), instead assume that f_1 and f_2 translate a single sense of e, and so are synonymous. Our work builds upon this observation, analyzing the various possible relations between senses and translations in greater detail, and leveraging them them to improve WSD.

Despite the early successes of translation-based WSD, methods based on monolingual resources, namely WordNet (Miller et al., 1990) and SemCor (Miller et al., 1993), became prominent in the 2010s. *It Makes Sense* (Zhong and Ng, 2010), a supervised WSD system based entirely on monolingual contextual features, remained state-of-the-art for most of the decade (Papandrea et al., 2017) before being replaced by methods based on contextual embeddings (Hadiwinoto et al., 2019). In the early 2020s, WSD systems leveraging increasingly sophisticated pre-trained language models approached and finally exceeded 80% accuracy on standard WSD datasets (Blevins and Zettlemoyer, 2020; Barba et al., 2021a,b). In response to these advances, Maru et al. (2022) proposed to focus on more difficult WSD instances, such as those involving rare senses, or on which modern WSD systems tend to make errors. We support this proposal, and make use of their "challenge" datasets in our experiments.

3 Mapping Senses and Translations

While the use of translation information to identify or even *define* word senses was frequent in early WSD research, today it primarily serves as supplementary data, rather than as the core of the method (Luan et al., 2020). In this section, we lay the theoretical groundwork for explaining this paradigm shift; an empirical analysis follows in the next section.

Given an ideal one-to-one mapping between senses of a word and its lexical translations, each sense could be unambiguously defined by a distinct translation, and each translation would indicate a different sense. Figure 1 shows a graphical representation of a sense-translation mapping which does *not* conform to this ideal, with three Italian translations of the English noun *wood*. An edge between a sense and a translation indicates that the former can be translated by the latter. As the sense-translation mapping is not bijective, we cannot use translation knowledge alone to determine the sense of an instance of *wood*.

We can analyze the theoretical properties of such a mapping in terms of three word-level binary predicates, which are defined on a given source word e and language of translation F. Each of these predicates is a necessary condition for such an ideal mapping to exist. Moreover, in conjunction, they represent a sufficient condition for using a word's translations as a sense inventory. The three sense-translation mapping predicates are discussed in the following subsections.

3.1 One Sense per Translation (OSPT)

One Sense per Translation (OSPT) is the key predicate for translation-based WSD, as it facilitates the inference of a word's sense from its translation. OSPT underlies the method that we propose in Section 5.

OSPT(e, F) := "all senses of the word e have disjoint sets of lexical translations in language F"

If OSPT holds, each translation of e corresponds to exactly one sense, and so we can use the sense-translation mapping to perform WSD. Exceptions to OSPT occur when words from different languages share multiple senses, a phenomenon which we refer to as *parallel polysemy*. For OSPT to hold, the source word cannot exhibit parallel polysemy with any of its translations. For example, Figure 1 shows a violation of OSPT, as the Italian word *legno* maps to two distinct senses of the English word *wood*. Therefore, the sense of *wood* in a given sentence cannot be inferred solely from the fact that it is translated as *legno*. On the other hand, if an instance of *wood* is translated into Italian as *selva*, we can infer that it is used in its "forest" sense.

3.2 One Translation per Sense (OTPS)

The One Translation per Sense (OTPS) predicate can be viewed as a dual of OSPT, reversing the roles of senses and translations.

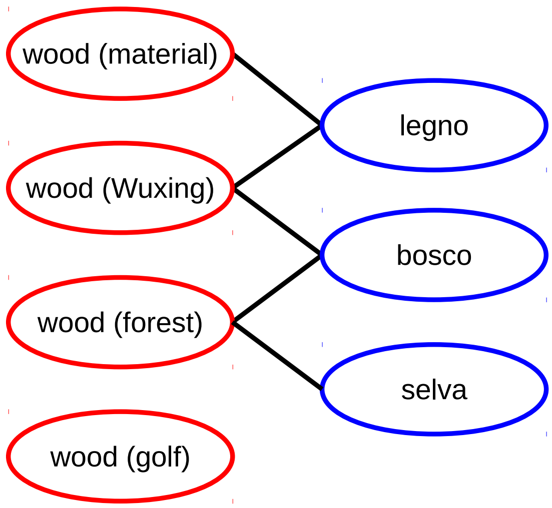

Figure 1: An example mapping between senses and translations. Each translation corresponds to at least one sense.

OTPS(e, F) := "each sense of the word e has at most one lexical translation in language F"

In other words, no pair of translations translate the same sense. Exceptions to OTPS are instances of synonymy between translations of a given source word.[1] For example, the "forest" sense in Figure 1 maps to two distinct translations, *bosco* and *selva*, violating OTPS. This presents a challenge to the proposal to use translations as sense inventories (Resnik and Yarowsky, 1997) by creating cases where instances of a word need not be distinguished by their translations. Moreover, this also poses a problem for aligning sense distinctions with translation distinctions (Lefever and Hoste, 2010), as *bosco* and *selva* must somehow be "clustered" to avoid identifying instances of *wood* with these translations as being semantically distinct. Note, however, that unlike violations of OSPT, translations that cause OTPS violations can still be used to disambiguate the translated word in some cases.

3.3 No Lexical Gaps (NoLG)

The No Lexical Gaps (NoLG) predicate reflects the importance of *lexical gaps* (Bentivogli and Pianta, 2000) in multilingual semantics.

NoLG(e, F) := "each sense of the word e has at least one translation in language F"

Since it is not practical to enumerate all possible

[1] Interestingly, the WSD algorithm of Diab and Resnik (2002), which disambiguates English words based on their French translations, is based on the assumption that all target-language words are monosemous.

phrasal translations of each sense, such lexical gaps generally preclude translation-based WSD: we cannot identify a sense based on its lexical translation if it *doesn't have* a lexical translation. For example, the "golf" sense of *wood* in Figure 1 corresponds to a lexical gap, and so would need to be translated into Italian by a compositional phrase, such as "legno da golf".

In summary, an ideal one-to-one sense-translation mapping seems to be a very brittle structure. Any exception to OSPT, OTPS, or NoLG would complicate the use of translations to define sense inventories. Moreover, any exception to OSPT or NoLG will outright preclude the use of translations alone for WSD. The viability of translation-informed WSD therefore rests on the extent to which these properties hold in practice, which we investigate in the next section.

4 Empirical Analysis

We focus on English, with three languages of translation which represent various degrees of relatedness to English: Italian, Polish, and Chinese, For each language, we compute the proportion of English words for which OSPT, OTPS, and NoLG hold in BabelNet (Navigli and Ponzetto, 2012), a large lexical knowledge base frequently used as a sense inventory for multilingual WSD (Pasini et al., 2021). We consider only English words with at least two senses in WordNet 3.0 (BabelNet inherits senses from WordNet), and at least one translation in the target language in BabelNet 4.0. There are 20,426 such words with Italian as the target language, 17,404 for Polish, and 19,973 for Chinese.

Table 1 summarizes the results. The NoLG values indicate that the majority of English words involve at least one lexical gap in any of the three languages of translation. The OTPS row shows that even fewer words have no more than one translation per sense. The OSPT property is more reliable, covering almost 60% words with Italian as the language of translation, and approaching 80% with less related languages such as Polish or Chinese. However, the last row in the table demonstrates that only a very small percentage of English words satisfy all three properties at the same time.

Since we have argued that the conjunction of the three properties is a necessary condition for an ideal one-to-one sense-translation mapping, these empirical results provide an explanation why using translations as sense inventories is infeasible

	Italian	Polish	Chinese
OSPT	59.5	77.4	75.7
OTPS	16.3	22.8	10.3
NoLG	47.4	38.3	40.3
ALL	1.9	2.6	1.5

Table 1: The percentage of English polysemous words in BabelNet which exhibit each of the three sense-translation mapping properties with respect to three languages of translation.

in practice. Furthermore, even if we had a sense inventory with a complete mapping between senses and translations (something BabelNet and comparable resources aspire to provide), the OSPT values in our results table indicate that a substantial portion of words cannot be disambiguated on the basis of their translations alone. We conclude that this was a key factor in the abandonment of the use of translations to induce sense inventories, or perform WSD on all words. Nevertheless, we posit that translations *can* be leveraged to improve WSD, specifically be exploiting those cases where a translation of a word in context uniquely determines its sense. In the next section, we present and apply a method for using translations to tag a subset of the tokens in a parallel corpus.

5 Corpus Tagging with OSPT

Although the results in Section 4 demonstrate that translations alone are not sufficient for all-words WSD, prior work such as Gale et al. (1992) and Lefever and Hoste (2010) have shown that they can still be applicable to lexical samples. In this section, we explore the idea of using translations to improve WSD on modern standard datasets. Specifically, we leverage those cases where the translation of a word corresponds to exactly one of its senses in order to create supplementary training data for a supervised WSD system.

5.1 Corpus Tagging

The generation of "silver datasets" for WSD is a way to address the knowledge acquisition bottleneck (Pasini, 2021), the difficulty of obtaining training data for supervised WSD. To this end, the goal of semantic corpus tagging is not to disambiguate all word tokens, or any particular subset of lemmas; rather, the goal is to partially sense-annotate a corpus to produce supplementary training data for a supervised WSD system.

Automatic sense tagging has been a popular area of research in lexical semantics. Taghipour and Ng (2015) used a mapping of Chinese translations to English senses to annotate the English side of an English-Chinese parallel corpus; however, this mapping is not available. Pasini and Navigli (2017) sense-tag Wikipedia articles using a variant of the personalized page-rank algorithm (PPR), while Delli Bovi et al. (2017) applies a similar approach to the EuroParl parallel corpus. Barba et al. (2020) use a pre-trained language model to identify semantically-equivalent translations of manually sense-annotated tokens. Most recently, Hauer et al. (2021) propose a family of pipeline approaches employing WSD methods, machine translation, lexical resources, and various filtering techniques.

Our work differs from prior work on using translations for WSD in that (a) we show that our method can achieve good results with only one language of translation, (b) our method is independent of statistical information such as relative sense frequencies, and (c) our method does not explicitly require any contextual information. In contrast, the method of Apidianaki and Gong (2015) backs off to the BabelNet first sense (BFS), a frequency-based baseline, if it is unable to narrow down the sense of the target word. This back-off strategy is particularly undesirable for tagging tokens that correspond to rare word senses. Moreover, their method is tested only with multiple languages of translation, and is applied directly to all-words WSD on a parallel corpus, rather than to generation of high-precision training data. The method of Bonansinga and Bond (2016) similarly depends on sense frequency information, and is evaluated only intrinsically, with multiple languages of translation. The method of Luan et al. (2020) depends on an existing disambiguation of the text, in addition to translations. Thus, our method is unique in that it can produce supplementary WSD training data with minimal assumptions about the available resources.

5.2 Method

Our method is inspired by Loureiro and Camacho-Collados (2020). They sense annotate only tokens that correspond to monosemous words, i.e., those that have only one sense, which is a trivial task in itself. However, they also show that a WSD method which propagates information between senses of different words can benefit from these annotations.

```
for each token e on the S side of C do
    if ∃ token f aligned with e then
        M_e ← the set of synsets containing e
        M_f ← the set of synsets containing f
        if |M_e ∩ M_f| = 1 then
            Let s be the sole synset in M_e ∩ M_f
            Tag e with sense (e, s)
```

Figure 2: Pseudo-code for the sense tagging algorithm.

For example, the monosemous word *airplane* is a synonym of the word *plane*, which is polysemous. Therefore, an annotated instance of *airplane* can inform a model about the context in which the corresponding sense of *plane* may appear.

In our approach, instead of monosemous words, we sense tag tokens which can be disambiguated based on their translations. For example, the English noun *vault* has four senses, corresponding to a burial vault, a bank vault, an arched ceiling, or a jump over an obstacle. The Polish word *wolta* can translate only the “jump” sense. Therefore, if we find an instance of *vault* translated as *wolta*, we can annotate *vault* with its “jump” sense, as no other sense could have been so translated. The absence of parallel polysemy between *vault* and *wolta* is a sufficient condition for the correctness of this annotation, regardless of whether OSPT holds for all Polish translations of *vault*. Our method uses this approach to partially annotate a parallel corpus, creating new sense-annotated WSD training data. Our hypothesis is that adding our translation-based annotations to a standard training corpus will improve the results of a supervised WSD system.

We follow the theoretical framework of Hauer and Kondrak (2023). The sense inventories, as well as the mapping between senses and translations, can be obtained from a multilingual wordnet, such as BabelNet. Multilingual wordnets consist of synonym sets, or *synsets*, each corresponding to a concept, and containing the words which can express that concept. The synsets that contain a word correspond to its senses; a sense can be viewed as a pair of a word and a synset that contains it. The target-language words in that synset are the words which can translate that sense. For example, in Figure 1, a multilingual wordnet should have a synset corresponding to the concept of “wood (material)” which contains *wood* and *legno*, but not *selva*.

The pseudo-code of the algorithm is shown in Figure 2. It takes as input a sentence-aligned parallel corpus C, involving the source language S (in our experiments, English) and the target language T, which has been tokenized, lemmatized, POS-tagged, and word-aligned. The algorithm generates sense tags for a subset of the tokens on the source side of C. The algorithm consults a wordnet that covers languages S and T. For each content word token e on the source side aligned with a single target-language token f, we determine the number of synsets which contain both e and f. Since each sense of a word uniquely corresponds to a synset containing that word, this is equivalent to determining how many senses of e can be translated by f. If the result is exactly one, we annotate e with its sense corresponding to the synset s that it shares with f. For example, if an instance of *wood* is aligned with *selva*, it is tagged with its "forest" sense, given that it is the only sense of *wood* which *selva* can translate.

Our method is unsupervised, efficient, scalable, and fully explainable. Its running time scales linearly with the size of the corpus. The resources upon which it depends are freely available for a wide variety of languages. These include the parallel corpora our method annotates, a multilingual wordnet, as well as tools for tokenization, POS-tagging, and alignment. It operates purely on the basis of contextual translation, without the need for additional tools such as knowledge-based WSD systems or contextual embeddings.

5.3 Intrinsic Evaluation

We test our translation-based corpus-tagging method on the manual sense annotations in MultiSemCor (MSC, Bentivogli and Pianta, 2005), a word-aligned sense-annotated bitext, which was created by manually translating SemCor (Miller et al., 1993). It is tokenized, POS-tagged, and word-aligned with a knowledge-based aligner. There are 91,937 English word tokens in MSC annotated with exactly one WordNet 1.6 sense, and aligned with a single Italian word. We randomly select 10,000 of these tokens, and strip them of their sense annotations to form our test set.

As our multilingual wordnet, we use MultiWordNet (MWN, Pianta et al., 2002) version 1.5.0. MWN was created by expanding Princeton WordNet 1.6 by adding Italian translations, as well as new synsets to cover English lexical gaps. To mitigate the sense omission errors in MWN, we enrich it with 81,937 sense-translation pairs from MSC, excluding those which are in our 10k-token test set.

The results of the application or our method to the 10,000 annotated tokens in the test set yield a coverage of 33.3% and a precision of 92.6%, with the majority of errors caused by missing translations in MWN. Thus, our unsupervised method achieves higher precision than contemporary supervised WSD systems on standard English WSD datasets (Barba et al., 2021b). While these results are not directly comparable due to the different test sets, we interpret this as strong evidence for the efficacy and utility of our method for generating high-quality WSD training data.

5.4 Extrinsic Evaluation

Having demonstrated that our method can accurately disambiguate a subset of the tokens in a corpus, in this section we test whether sense-annotated data produced in this way can be used to improve the performance of a supervised WSD system. This is achieved by appending the data that our translation-based method produces to SemCor, a standard training corpus for English WSD. Note that no manual sense annotations exist for the corpus that we annotate in these experiments; we are creating novel sense-annotated data.

5.4.1 Experimental Setup

Our parallel corpus is the English-Italian part of the OpenSubtitles corpus (Lison and Tiedemann, 2016), which contains approximately 35M sentence pairs. We tokenize, lemmatize, and POS-tag both sides of the corpus with TreeTagger (Schmid, 2013) using pre-trained models.[2] We perform word alignment with BabAlign (Luan et al., 2020), which refines the output of FastAlign (Dyer et al., 2013) by leveraging BabelNet as a source of lexical knowledge.

We again derive a sense-translation mapping from MultiWordNet, but this time without adding information from MultiSemCor. Since MultiWordNet is based on WordNet 1.6, we map each sense annotation to its most probable WordNet 3.0 equivalent, using a publicly available probabilistic mapping.[3]

As our supervised WSD system, we adopt the latest version of LMMS (Loureiro et al., 2022), which exploits relations between senses derived from WordNet in order to share information across related senses.

[2] https://cis.uni-muenchen.de/~schmid
[3] http://www.lsi.upc.es/~nlp

Corpus	Tokens	Senses	Lemmas
SemCor	226,036	33,316	22,899
F10	219,793	28,589	23,033
FFSC	117,646	16,818	15,329
FFLC	90,616	13,147	12,406

Table 2: Statistics on the sets of sense annotations generated using the three filtering procedures.

Dataset	Full	MFS	LFS	ZSS	ZSL
SE2	2,282	1,486	796	385	255
SE3	1,850	1,213	637	198	112
S07	455	250	205	53	20
S13	1,644	1,031	613	341	202
S15	1,022	623	399	204	103
ALL	7,253	4,603	2,650	1,181	692

Table 3: Number of instances in each of the subsets of each dataset and the concatenation of all five datasets.

5.4.2 Filtering Annotations

Supervised WSD systems tend to exhibit a bias toward senses which are more frequent in the training data (Loureiro et al., 2020). Therefore, even a set of perfectly correct sense annotations may degrade the model's performance if the sense frequency distribution in the newly produced data diverges from that of the test data, which is not known in advance. We therefore filter the generated annotations to avoid greatly altering the sense frequency distribution of SemCor.

Following the example of Loureiro and Camacho-Collados (2020), we limit the number of annotated instances of each individual sense to 10, selected at random. This not only helps to prevent highly unbalanced sense frequency distributions, but also reduces the training time on the generated corpora. We refer to this set of instances as `F10`. In order to focus on gaps in the coverage on SemCor, we also test two additional filtering strategies that are applied to the annotations in `F10`. The first filters for lemma coverage (`FFLC`), by removing all annotations for *lemmas* which appear in SemCor. The second filters for sense coverage (`FFSC`), by removing all annotations for *senses* which appear in SemCor. Therefore, the `FFLC` annotations are a subset of the `FFSC` annotations, which in turn are a subset of the `F10` annotations.

5.4.3 Datasets

We obtain baseline results by training LMMS on SemCor, specifically the version provided by Raganato et al. (2017). To test our method, we train three additional LMMS models which augment SemCor annotations with `F10`, `FFSC`, and `FFLC`, respectively. The sizes of these generated supplementary datasets, and of SemCor itself, are shown in Table 2.

We evaluate our models on the standard WSD benchmark of Raganato et al. (2017, "R17"). In addition to providing the baseline SemCor training corpus, R17 also contains five English WSD test sets created for five shared tasks: Senseval-2 (SE2, Edmonds and Cotton, 2001), Senseval-3 (SE3, Snyder and Palmer, 2004), SemEval-2007 (S07, Pradhan et al., 2007), SemEval-2013 (S13, Navigli et al., 2013), and SemEval-2015 (S15, Moro and Navigli, 2015). Following prior work, we use the S07 dataset to develop our method. We also evaluate our models on the concatenation of all five datasets, referred to as ALL[4], using the provided evaluation program; since LMMS disambiguates all words, the metrics precision, recall, F1, and accuracy are all equal throughout these experiments.

Following Blevins and Zettlemoyer (2020), we also evaluate our models on the following subsets of ALL: most frequent sense (MFS), less frequent sense (LFS), zero-shot senses (ZSS), and zero-shot lemmas (ZSL). MFS and LFS are disjoint, and their union is the complete dataset; ZSL is a subset of ZSS. Table 3 shows the size of each such subset.

Finally, we also test our models on five new benchmark datasets of Maru et al. (2022, "M22"): challenge (42D), amended ALL (ALLa), amended S10 (S10a), hardEN(hEN), and softEN (sEN).

5.4.4 Results

The results in Tables 4 and 5 show that adding supplementary training data created by our method generally increases WSD accuracy, especially on rare and unseen senses. On the recently proposed 42D and hardEN challenge sets, we observe accuracy improvements of 4.6% and 2.5% respectively, using the `F10` filtering strategy. This same approach yields improvements on LFS, ZSS, and ZSL partitions of the R17 ALL set, demonstrating that our method makes models more robust against such instances. We interpret these results as evidence for the efficacy and utility of our translation-based corpus tagging method.

The results further suggest that filtering generated annotations has a substantial impact on the resulting model. The frequency with which a word can be tagged with a particular sense by leveraging

[4]This includes S07, as is standard in the WSD literature.

Training Data	R17						M22				
	SE2	SE3	S07	S13	S15	ALL	42D	ALLa	S10a	hEN	sEN
SemCor (Baseline)	76.1	**73.9**	67.0	**75.2**	77.4	75.0	35.9	74.9	**77.3**	12.6	**78.0**
SemCor + `F10`	74.9	72.6	65.9	72.8	78.2	73.7	**40.5**	73.3	77.1	**15.1**	76.6
SemCor + `FFLC`	**76.7**	**73.9**	**67.5**	75.0	77.5	**75.1**	34.9	**75.1**	76.6	13.4	77.9
SemCor + `FFSC`	76.2	72.3	66.8	73.5	**78.3**	74.3	38.4	74.1	76.3	14.7	77.1

Table 4: F1-scores (in %) on the 10 WSD test sets. SE07 is the development set. The best results are in bold.

Training Data	R17 - ALL			
	MFS	LFS	ZSS	ZSL
SemCor (Baseline)	85.4	51.2	58.9	88.9
SemCor + `F10`	83.1	**52.1**	61.7	89.5
SemCor + `FFLC`	**85.5**	51.3	60.1	89.6
SemCor + `FFSC`	83.9	51.9	**62.7**	**89.7**

Table 5: F1-scores (in %) on subsets of the concatenation of all R17 datasets. The best results are in bold.

lexical translation need not correlate with the frequency of that sense in practice. Therefore, when using such generated corpora, care should be taken to select an appropriate filtering strategy. For instance, in a corpus where unseen senses or words are expected (e.g., in an unusual genre or domain), the `FFSC` filtering strategy may be the best option, as shown by its accuracy yields on ZSS and ZSL instances.

We conclude that our method for translation-based sense tagging offers substantial benefits, especially on difficult instances (Blevins et al., 2021). These improvements are obtained using a recent WSD method which is based on pre-trained transformer-based language models. This demonstrates that lexical translation can be a useful source of information even for modern WSD systems.

As a final note, we note that since the phenomenon of *parallel polysemy* is closely related to that of *parallel homonymy*, our approach is well-suited to homonym-level disambiguation. Hauer and Kondrak (2020) argue that homonym distinctions are the coarsest possible sense inventory, and that almost all homonyms have disjoint sets of translations. Therefore, unlike OSPT, One Homonym per Translation (OHPT) *does* hold in general. Our translation-based approach could therefore be applied with near-perfect accuracy to disambiguate words at the homonym level.

6 Discussion

Our theoretical analysis in Section 3 established that OSPT is a sufficient condition for the ability to determine the sense of a word given its translation in context. However, the subsequent empirical analysis in Section 4 showed that OSPT does not hold in general. Nevertheless, our experiments in Section 5 provide clear evidence that we can leverage translations to produce high-precision sense annotations on the subset of word instances for which OSPT holds. These results demonstrate the importance of investigating the relations between senses, synonymy, polysemy, and translation. In this section, we further explore these ideas, taking the assumptions examined by Yao et al. (2012) (c.f., Section 2) to their logical extremes.

6.1 One Concept per Word: No Polysemy

First, let us consider a hypothetical language in which polysemy does not exist; that is, every content word has exactly one sense. In such a language, there could be no semantic ambiguity, and so WSD would be trivial: any given word could only express a single concept, regardless of its context. OSPT would always hold in such a language, no matter the language of translation, since each translation of a word could only translate its single sense.

To the best of our knowledge, no natural language contains only monosemous words. For example, 77.8% of English words in BabelNet occur in only one synset, with many of those being rare or technical terms. Similarly, Loureiro and Camacho-Collados (2020) observe that nearly 80% of lemmas in WordNet have only one sense, which allows them to generate useful resources for WSD. Only some constructed languages, such as Lojban/Loglan, strive to enforce complete monosemy on the lexicon (Cowan, 1997).

The untenable position that rejects any partitioning of word meanings into senses ("one sense per word") relates to various approaches to both theoretical and computational linguistics. In theoretical linguistics, the *monosemist* approach holds that different observed senses of a polysemous word result from a combination of its unique core meaning with the pragmatics of each specific context (François, 2008). In computational linguistics, methods that rely on exclusive use of static word embeddings, such as those learned by word2vec (Mikolov et al.,

2013) make no allowance for discrete senses or sense embeddings.

6.2 One Word per Concept: No Synonymy

Now, let us consider the opposite extreme: a hypothetical language without synonymy. If a wordnet were constructed for such a language, every synset would contain exactly one word. For any given concept, there would be at most one word that could be used to express it. One Translation per Sense (OTPS) would always hold if such a language was used as the language of translation.

Again, it is unlikely that the entire lexicon of any natural language could satisfy this requirement. A language could perhaps be *constructed* according to this principle: for example, in Esperanto, synonymy and homonymy are considered undesirable (Puškar, 2015). Moreover, there will be a subset of any language which *does* satisfy this property. Indeed, approximately 56% of WordNet 3.0 synsets contain only one word (e.g., *proton*).

A similar position in computational linguistics ("one sense per context") is diametrically opposite to the monosemist approach described above. For example, Martelli et al. (2021) propose "dropping the requirement of a fixed sense inventory" and instead using representations which assign each word token a unique contextualized embedding. Such a position can be interpreted as an assignment of a unique sense to every occurrence of a given word in a distinct context. In view of our theoretical investigation, such an approach is effectively incompatible with our definition of synonymy. Nevertheless, the existence of synonymy in any human language is widely accepted in linguistics. In addition, computational linguistics tasks, such as machine translation, need to account for synonymy, given that the goal is to produce fully fluent, rather than just semantically correct texts and utterances.

6.3 One Word ≡ One Concept

If the two constraints described above are combined, it would result in a language that has neither polysemy nor synonymy. We refer to this hypothetical language as *Interlingua*. In Interlingua, every concept could be expressed by exactly one word, which could express only that concept; every synset would have a size of one, and every word would be in one synset. Assuming a sense-translation mapping is available, e.g. via a multilingual wordnet which includes Interlingua, lexical translation *into* Interlingua could be reduced to identifying the sense of the source word. The converse also holds: the sense of a word could always be identified, given its translation into Interlingua. Working in the other direction, given a perfect multilingual wordnet, finding a translation for an Interlingua word would only require selecting a word from the corresponding synset in the target language.

Perhaps the most direct application for Interlingua is language-independent semantic parsing. Martínez Lorenzo et al. (2022) propose the *BabelNet Meaning Representation* (BMR), a semantic parsing formalism which converts an input sentence into a language-independent representation. Each content word is mapped to the unique identifier of the BabelNet synset corresponding to the concept it refers to. This creates a formal metalanguage in which every concept is unambiguously expressed in exactly one way: by the corresponding BabelNet synset ID. Hence, the BMR satisfies one "word" per concept *and* one concept per "word", with BabelNet IDs taking the place of words. There is no synonymy, as each ID is by design unique in representing its particular concept, nor is there polysemy, as each ID is unambiguous in its reference to some lexicalized concept. Thus, what may appear as a completely hypothetical and abstract construct can in fact be viewed as a theoretical model of a modern semantic approach.

7 Conclusion

In this paper, we formulate several propositions related to senses, translations, synonymy, and polysemy. We show empirically that the assumptions that would allow translations to serve as a sense inventory hold simultaneously only for a small fraction of words. Nevertheless, we also demonstrate that the link between word senses and translations is not merely of theoretical interest. In particular, we present a method for leveraging translations to perform high-precision unsupervised sense annotation. We observe substantial WSD improvements especially on senses or lemmas that are less frequent or not found at all in existing training data.

Considering the above applications to constructed languages, contextual embeddings, and semantic parsing, we intend to continue our theoretical investigations into open issues in multilingual lexical semantics, and guide empirical research toward more explainable models and results.

Limitations

The principal limitation of our sense-tagging method is its dependence on linguistic resources, particularly text corpora and multilingual wordnets. As is the case with any method which depends on such resources, the reliability of our method will vary depending on the language to which it is applied, and the quality of the resources available. Care should be taken when applying our method to languages and domains, where resources are limited in terms of availability, coverage, or correctness. Any biases in these resources, e.g. biases toward English, may be inherited by our method. Likewise, the quality of translation and word alignment methods for pairs of languages will have a substantial impact on the quality of the data our method produces. Thus, before applying our method, we recommend assessing the quality of semantic resource coverage and translation and alignment quality for the languages under consideration. Nevertheless, the state of resource coverage and quality within NLP is improving, and we expect the applicability of our method to improve concordantly.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Alberta Machine Intelligence Institute (Amii). We also thank Robert Holte for providing additional comments.

References

Marianna Apidianaki. 2008. Translation-oriented word sense induction based on parallel corpora. In *Language Resources and Evaluation (LREC)*.

Marianna Apidianaki and Li Gong. 2015. LIMSI: Translations as source of indirect supervision for multilingual all-words sense disambiguation and entity linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 298–302.

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 597–604.

Edoardo Barba, Tommaso Pasini, and Roberto Navigli. 2021a. ESC: Redesigning WSD with extractive sense comprehension. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4661–4672.

Edoardo Barba, Luigi Procopio, Niccolo Campolungo, Tommaso Pasini, and Roberto Navigli. 2020. MuLaN: Multilingual label propagation for word sense disambiguation. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3837–3844.

Edoardo Barba, Luigi Procopio, and Roberto Navigli. 2021b. ConSeC: Word sense disambiguation as continuous sense comprehension. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1492–1503, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Luisa Bentivogli and Emanuele Pianta. 2000. Looking for lexical gaps. In *Proceedings of the ninth EURALEX International Congress*, pages 8–12. Stuttgart: Universität Stuttgart.

Luisa Bentivogli and Emanuele Pianta. 2005. Exploiting parallel texts in the creation of multilingual semantically annotated resources: The MultiSemCor Corpus. *Natural Language Engineering*, 11(3):247–261.

Terra Blevins, Mandar Joshi, and Luke Zettlemoyer. 2021. FEWS: Large-scale, low-shot word sense disambiguation with the dictionary. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 455–465.

Terra Blevins and Luke Zettlemoyer. 2020. Moving down the long tail of word sense disambiguation with gloss informed bi-encoders. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1006–1017.

Giulia Bonansinga and Francis Bond. 2016. Multilingual sense intersection in a parallel corpus with diverse language families. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 44–49.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991. Word-sense disambiguation using statistical methods. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 264–270.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 33–40.

John Woldemar Cowan. 1997. *The complete Lojban language*, volume 15. Logical Language Group.

Ido Dagan, Alon Itai, and Ulrike Schwall. 1991. Two languages are more informative than one. In *29th Annual Meeting of the Association for Computational Linguistics*, pages 130–137.

Claudio Delli Bovi, Jose Camacho-Collados, Alessandro Raganato, and Roberto Navigli. 2017. Eurosense: Automatic harvesting of multilingual sense annotations from parallel text. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 594–600.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Mona Diab and Philip Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 255–262.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648.

Philip Edmonds and Scott Cotton. 2001. Senseval-2: Overview. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 1–5.

Alexandre François. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, 163-215.

William A Gale, Kenneth W Church, and David Yarowsky. 1992. Using bilingual materials to develop word sense disambiguation methods. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, volume 112. Citeseer.

Christian Hadiwinoto, Hwee Tou Ng, and Wee Chung Gan. 2019. Improved word sense disambiguation using pre-trained contextualized word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5297–5306, Hong Kong, China. Association for Computational Linguistics.

Bradley Hauer and Grzegorz Kondrak. 2020. One homonym per translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7895–7902.

Bradley Hauer and Grzegorz Kondrak. 2023. Synonymy = translational equivalence. *arXiv preprint arXiv:2004.13886*.

Bradley Hauer, Grzegorz Kondrak, Yixing Luan, Arnob Mallik, and Lili Mou. 2021. Semi-supervised and unsupervised sense annotation via translations. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 504–513, Held Online. INCOMA Ltd.

Nancy Ide. 2000. Cross-lingual sense determination: Can it work? *Computers and the Humanities*, 34(1-2):223–234.

Els Lefever and Veronique Hoste. 2010. SemEval-2010 task 3: Cross-lingual word sense disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 15–20.

Els Lefever, Véronique Hoste, and Martine De Cock. 2011. ParaSense or how to use parallel corpora for word sense disambiguation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 317–322.

Pierre Lison and Jörg Tiedemann. 2016. OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016*, pages 923–929. European Language Resources Association.

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.

Daniel Loureiro and Jose Camacho-Collados. 2020. Don't neglect the obvious: On the role of unambiguous words in word sense disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3514–3520.

Daniel Loureiro, Alípio Mário Jorge, and Jose Camacho-Collados. 2022. LMMS reloaded: Transformer-based sense embeddings for disambiguation and beyond. *Artificial Intelligence*, 305.

Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. 2020. Analysis and evaluation of language models for word sense disambiguation.

Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. 2021. Analysis and evaluation of language models for word sense disambiguation. *Computational Linguistics*, 47(2):387–443.

Yixing Luan, Bradley Hauer, Lili Mou, and Grzegorz Kondrak. 2020. Improving word sense disambiguation with translations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4055–4065.

Federico Martelli, Najla Kalach, Gabriele Tola, and Roberto Navigli. 2021. SemEval-2021 Task 2: Multilingual and Cross-lingual Word-in-Context Disambiguation (MCL-WiC). In *Proceedings of the Fifteenth Workshop on Semantic Evaluation (SemEval-2021)*.

Abelardo Carlos Martínez Lorenzo, Marco Maru, and Roberto Navigli. 2022. Fully-Semantic Parsing and Generation: the BabelNet Meaning Representation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1727–1741, Dublin, Ireland. Association for Computational Linguistics.

Marco Maru, Simone Conia, Michele Bevilacqua, and Roberto Navigli. 2022. Nibbling at the hard core of Word Sense Disambiguation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4724–4737, Dublin, Ireland. Association for Computational Linguistics.

Rowan Hall Maudslay and Simone Teufel. 2022. Metaphorical polysemy detection: Conventional metaphor meets word sense disambiguation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 65–77, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Rada Mihalcea, Ravi Sinha, and Diana McCarthy. 2010. SemEval-2010 task 2: Cross-lingual lexical substitution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 9–14, Uppsala, Sweden.

Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*. Technical report.

George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. 1990. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244.

George A. Miller, Claudia Leacock, Randee I. Tengi, and Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of the ARPA Workshop on Human Language Technology*, pages 303–308.

Andrea Moro and Roberto Navigli. 2015. Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 288–297.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. Semeval-2013 task 12: Multilingual word sense disambiguation. In *Proceedings of the Seventh International Workshop on Semantic Evaluation*, pages 222–231.

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.

Simone Papandrea, Alessandro Raganato, and Claudio Delli Bovi. 2017. SupWSD: A flexible toolkit for supervised word sense disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 103–108, Copenhagen, Denmark. Association for Computational Linguistics.

Tommaso Pasini. 2021. The knowledge acquisition bottleneck problem in multilingual word sense disambiguation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4936–4942.

Tommaso Pasini and Roberto Navigli. 2017. Train-o-matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 78–88.

Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021. Xl-wsd: An extra-large and cross-lingual evaluation framework for word sense disambiguation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13648–13656.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. Multiwordnet: developing an aligned multilingual database. In *First international conference on global WordNet*, pages 293–302.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations*, pages 87–92.

Krunoslav Puškar. 2015. Esperanto (s) en perspektivo? Croatian esperantists on the international language Esperanto. *Interdisciplinary Description of Complex Systems: INDECS*, 13(2):322–341.

Alessandro Raganato, Iacer Calixto, Jose Camacho-Collados, Asahi Ushio, and Mohammad Taher Pilehvar. 2023. SemEval-2023 Task 1: Visual Word Sense Disambiguation (Visual-WSD)). In *FORTHCOMING: Proceedings of the Seventeenth Workshop on Semantic Evaluation (SemEval-2023)*.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 99–110.

Philip Resnik and David Yarowsky. 1997. A perspective on word sense disambiguation methods and their evaluation. In *Tagging Text with Lexical Semantics: Why, What, and How?*, pages 79–86.

Helmut Schmid. 2013. Probabilistic part-of-speech tagging using decision trees. In *New methods in language processing*, page 154.

Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational linguistics*, 24(1):97–123.

Benjamin Snyder and Martha Palmer. 2004. The English all-words task. In *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 41–43.

Kaveh Taghipour and Hwee Tou Ng. 2015. One million sense-tagged instances for word sense disambiguation and induction. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 338–344.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Warren Weaver. 1949. Translation. *Machine Translation of Languages*.

Xuchen Yao, Benjamin Van Durme, and Chris Callison-Burch. 2012. Expectations of word sense in parallel corpora. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 621–625.

David Yarowsky. 1995. Unsupervised WSD rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, Massachusetts Institute of Technology, Cambridge, MA*.

Zhi Zhong and Hwee Tou Ng. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations*, pages 78–83.