# One Homonym per Translation

**Bradley Hauer, Grzegorz Kondrak**

Department of Computing Science
University of Alberta, Edmonton, Canada
{bmhauer, gkondrak}@ualberta.ca

## Abstract

The study of homonymy is vital to resolving fundamental problems in lexical semantics. In this paper, we propose four hypotheses that characterize the unique behavior of homonyms in the context of translations, discourses, collocations, and sense clusters. We present a new annotated homonym resource that allows us to test our hypotheses on existing WSD resources. The results of the experiments provide strong empirical evidence for the hypotheses. This study represents a step towards a computational method for distinguishing between homonymy and polysemy, and constructing a definitive inventory of coarse-grained senses.

## 1 Introduction

Many words are semantically ambiguous, in that they have multiple senses. The relationship between two senses of a word is called *polysemy* if they are semantically related, and *homonymy* otherwise (Jurafsky and Martin 2009). Senses that belong to the same homonym are polysemous (e.g. #2 and #5 in Table 1), while senses of distinct homonyms are homonymous (e.g. #2 and #1 in Table 1).

The differentiation of homonymous and polysemous word senses is one of the central problems of lexicography (Mel'čuk 2013). A textbook on thoeretical semantics devotes an entire chapter to the problem, concluding that it may be insoluble, as the intuitions of native speakers cannot be relied upon (Lyons 1995). Psycho-linguistics furnishes evidence for a common representation of closely related senses in the mental lexicon (Brown 2008), which suggests that NLP applications would benefit from the ability to distinguish homonym-level meaning differences (Utt and Padó 2011). In fact, standard NMT systems make a substantial number of errors on homonyms (Liu, Lu, and Neubig 2018).

The study of homonymy is also of utmost importance to the problem of establishing the set of senses for a given word. In word sense disambiguation (WSD), which is the task of selecting the intended sense of an ambiguous word token, the quality and granularity of the sense inventory greatly influences the design, evaluation, and utility of any system. The standard sense inventory, WordNet (Fellbaum 1998), makes no distinction between homonymy

| BANK$_n^1$ | | BANK$_n^2$ | |
|---|---|---|---|
| #2 | financial institution | #1 | sloping land |
| #5 | stock held in reserve | #3 | long ridge or pile |
| #6 | funds held by a house | #4 | arrangement of objects |
| #8 | container for money | #7 | slope in a road |
| #9 | building | #10 | flight maneuver |

Table 1: The senses of the noun 'bank' from WordNet 3.0, grouped by its two homonyms.

and polysemy, and is widely considered to be excessively fine-grained for many practical applications (Navigli 2018), as evidenced by a low inter-annotator agreement (Snyder and Palmer 2004). This has inspired substantial prior work on clustering fine-grained senses to create more coarse-grained sense inventories (Hovy et al. 2006; Navigli 2006; Snow et al. 2007; Dandala et al. 2013; McCarthy, Apidianaki, and Erk 2016).

Following the observation that different senses of a word often correspond to distinct words in another language (Resnik and Yarowsky 1997), another branch of prior work has sought to use translations to define sense inventories (Resnik and Yarowsky 1999; Diab and Resnik 2002; Ng, Wang, and Chan 2003; Chan, Ng, and Chiang 2007; Apidianaki 2008; Bansal, DeNero, and Lin 2012; Taghipour and Ng 2015). In order to be successful, such an approach would have to resolve the challenging issues of mapping senses to translations in a set of diverse target languages, as well as projecting them onto a standard sense inventory, such as WordNet.

In summary, clustering fine-grained senses and defining sense distinctions using translations are two competing methodologies for creating coarse-grained sense inventories. Regardless of which one is adopted, an understanding of the nature and characteristics of homonymous senses is a necessary step toward a principled method of defining senses and sense distinctions. In particular, distinctions between homonymous senses must be preserved in any sense inventory. This motivates our study, which contributes to such an understanding by directly linking homonymy to the concepts of translation and sense clustering, and thus bridging the gap between the two approaches.

The contributions of this work are both theoretical and empirical. The main goal is to create theoretical foundations

for the study of homonymy, which could pave the way for developing a computational method for distinguishing between homonymy and polysemy, and facilitate the task of constructing a definitive inventory of coarse-grained senses. We propose four hypotheses about the unique behavior of homonyms in the context of translations, discourses, collocations, and sense clusters. The hypotheses are formulated using established semantic concepts, and formalized in mathematical notation. Our principal hypothesis, as stated in the title, implies a sufficient condition for polysemy which is observable and replicable.

Apart from introducing the hypotheses, we perform experiments to provide empirical evidence for them. It is clear from prior work that what is true at one level of semantic granularity may not be true at another. For example, the well-known hypotheses *one sense per discourse* and *one sense per collocation* have been found not to hold consistently for WordNet senses. It is critical that all claims be formally stated and experimentally tested, regardless of whether the results are considered surprising; we have found no prior work that fulfills this requirement with respect to the four hypotheses presented in this paper. To facilitate our experiments, we create a new annotated resource, by identifying nearly three thousand English homonyms, and mapping them onto WordNet senses. The results of our experiments on multiple annotated corpora and language pairs strongly support our hypotheses.

## 2 Homonym Hypotheses

In this section, we formally define the notion of a homonym, and formulate our hypotheses using set notation. We attempt to keep the notational complexity to a minimum, while at the same time striving to avoid ambiguity.

### 2.1 Preliminaries

*Lexemes* are units of language that are represented in the lexicon (Murphy and Koskela 2010). *Words* are sets of word-forms that represent lexemes, and are associated with certain morpho-syntactic properties. This definition of words includes compounds, such as 'single out', as is the case in WordNet. We consider both lexemes and words that differ in part of speech as distinct. We write lexemes in capital letters, abstract words in single quotes, actual word-forms in italics, and sense meanings in double quotes. For example, the lexeme $\text{CUT}_v$ is represented by the verb 'cut', with the word-forms *cut, cuts,* and *cutting*. A lexeme is called polysemous if it contains multiple senses, and monosemous if it has only a single sense. Senses that belong to the same lexeme are semantically related, and therefore polysemous (Jurafsky and Martin 2009).

A *homonymous word* (e.g., the noun 'bank' in Table 1) represents more than one lexeme, and those lexemes are called *homonyms*. Senses associated with distinct homonyms are unrelated and therefore homonymous (Murphy and Koskela 2010). Consequently, the problem of deciding whether two senses of a homonymous word are polysemous is equivalent to deciding whether they belong to the same lexeme. Furthermore, since a non-homonymous word represents only a single lexeme, all of its senses are polysemous.

We are now ready to formally define homonyms. Let $\mathcal{L}$ and $\mathcal{W}$ denote the sets of lexemes and words of a given language, respectively, and let $w\colon \mathcal{L} \mapsto \mathcal{W}$ be a function that maps each lexeme to the word that represents it. In later sections, we will use $w^{-1}\colon \mathcal{W} \mapsto \mathcal{P}(\mathcal{L})$ to denote the function which maps each word to the set of lexemes it represents. We define the set of homonymous words $\mathcal{H}$ as the set of all words that represent multiple lexemes:

$$\mathcal{H} \stackrel{\text{def}}{=} \{W \in \mathcal{W} \mid \exists L, L' \in \mathcal{L} :$$
$$(L \neq L') \wedge (w(L) = w(L') = W)\}$$

For example, $w(\text{BANK}_n^1) = w(\text{BANK}_n^2) = $ 'bank' $\in \mathcal{H}$.

### 2.2 One Homonym per Translation

In general, there is no simple correspondence between word senses and their translations: a single sense may be translated by any of several synonyms, and different senses of the same word may have the same translation. (Ide and Wilks 2007) observe that cross-lingual distinctions often correspond to homonym-level disambiguation. We posit a direct relationship between translations and homonyms. Intuitively, if we randomly selected two different words from a bilingual dictionary, we would not expect them to have translations in common. The same reasoning applies to homonyms, since they are semantically unrelated lexemes that coincidentally share the same form. We formalize this insight as our principal hypothesis.

Put simply, the one homonym per translation hypothesis (OHPT) states that homonyms have disjoint translation sets. Formally, let $T(L)$ be a set of translations of a lexeme $L$, and let $w^{-1}$ be as defined as in Section 2.1. Then,

$$\forall H \in \mathcal{H} : \forall L, L' \in w^{-1}(H) :$$
$$(L \neq L') \Rightarrow T(L) \cap T(L') = \emptyset$$

For example, the Italian translations of the noun 'yard' can be partitioned into two disjoint sets $T(\text{YARD}_n^1) = $ {'iarda','yard'} and $T(\text{YARD}_n^2) = $ {'cortile','giardino'}, which correspond to two English homonyms, with the meanings of "unit" and "garden", respectively.

This hypothesis implies an important generalization: *the existence of a shared translation is a sufficient condition for polysemy*. Indeed, for homonymous words, senses that can be translated by the same word must belong to the same lexeme, and so are polysemous. As all other words represent only single lexemes, all their senses are polysemous by definition (Section 2.1). Therefore, we consider the OHPT hypothesis as a major step towards solving the problem of distinguishing between homonymy and polysemy.

### 2.3 One Homonym per Discourse

The *one sense per discourse* (OSPD) hypothesis was introduced in the seminal paper of (Gale, Church, and Yarowsky 1992). They observe that "well-written discourses tend to avoid multiple senses of a polysemous word", and confirm that the property holds with high probability on a set of
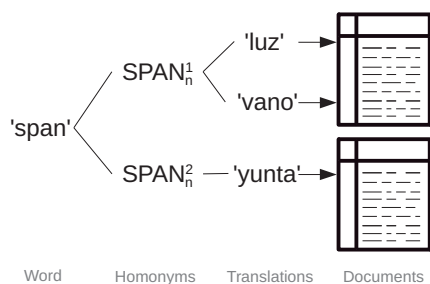
Figure 1: An example of an exception to the *one translation per discourse* hypothesis of (Carpuat 2009). The top two Spanish translations of 'span' are synonymous.

82 instance pairs involving 9 ambiguous words. However, (Krovetz 1998) reports that OSPD holds for only 67% of ambiguous words in SemCor, and conjectures that the hypothesis may only apply to homonymous senses.

We formulate Krovetz's conjecture as the *one homonym per discourse* hypothesis (OHPD), which can be viewed as a specialization of OSPD to homonyms. The hypothesis states that *all occurrences of a homonymous word in a discourse represent the same homonym*. A possible explanation of this phenomenon is that writers avoid the use of homonyms by employing their synonyms in order to reduce ambiguity in a discourse. Another explanation is that most discourses cover topics within a single domain, and therefore are unlikely to contain lexemes that are completely unrelated to each other.

Our formulation of the OHPD hypothesis states that no more than one lexeme of a homonymous word occurs in any given discourse. Formally, let $D$ be the set of lexemes that occur in a discourse, and let $w$ be again the function that maps lexemes to words. Then,

$$\forall L, L' \in D : (w(L) = w(L')) \Rightarrow (L = L')$$

We close this section by considering the relationship between OHPD and the *one translation per discourse* (OTPD) hypothesis of (Carpuat 2009). They report that approximately 80% of French words have a single English translation per document, which they interpret as strong support for their hypothesis. We note that the conjunction of our OHPT and OHPD hypotheses does not imply OTPD. Indeed, consider the example in Figure 1, which shows how the occurrence of three Spanish translations of the homonymous noun 'span' in two different documents leads to a violation of OTPD, but not of OHPD or OHPT.

### 2.4 One Homonym per Collocation

(Yarowsky 1993) proposes the *one sense per collocation* (OSPC) hypothesis, broadly defining a collocation as "the co-occurrence of two words in some defined relationship". Yarowsky reports that the hypothesis holds with the average 95% precision on a sample of words of an unreported size. However, (Martinez and Agirre 2000) find much weaker evidence for OSPC on WordNet senses, with precision values rarely exceeding 70%.

The explicit focus of (Yarowsky 1993) is on the most coarse-grained sense distinctions. Their word sample includes pseudo-words, words with different French translations, words spelled the same but pronounced differently (*homographs*), words pronounced the same but spelled differently (*homophones*), and words that are visually confusable in optical character recognition. All these types of words can be viewed as approximations of homonymy, as they involve pairs of distinct lexemes. We formalize this notion with the *one homonym per collocation* (OHPC) hypothesis, which states that only one homonym of a word should appear in any given collocation.

Formally, given a corpus of text, let $\mathcal{R}$ be the set of all collocations. For lexeme $L \in \mathcal{L}$, and collocation $r \in \mathcal{R}$, let $C_r(L)$ be a proposition which is true if and only if $w(L)$ occurs in collocation $r$ in the corpus. Then,

$$\forall H \in \mathcal{H} : \forall L, L' \in w^{-1}(H) : \forall r \in \mathcal{R} :$$
$$(C_r(L) \wedge C_r(L')) \Rightarrow (L = L')$$

For example, if $BANK_n^1$ ("repository") is found to occur in the collocation [word-to-right = *hired*] then $BANK_n^2$ ("ridge") is unlikely to occur in this collocation.

### 2.5 One Homonym per Sense Cluster

*Sense clustering* is the task of grouping together senses that are closely related (Dandala et al. 2013). Although the criteria for eliminating sense distinctions vary depending on the purpose of the sense inventory, a common motivation is to reduce the excessive granularity of WordNet (Snow et al. 2007). In particular, a manual clustering of WordNet senses was created as part of the OntoNotes project, with the objective of increasing the inter-annotator agreement on WSD to 90% (Hovy et al. 2006). Sense clustering has been shown to improve performance on a number of NLP tasks (Pilehvar et al. 2017), and can serve as an extrinsic evaluation for learned representations of senses (Mancini et al. 2017).

Since homonyms are distinct lexemes, we posit that any well-grounded clustering approach must avoid merging homonymous senses. Formally, let $\mathcal{C}$ be a sense clustering, a set of disjoint sets of senses, and let $S(L)$ be the set of senses of lexeme $L$. Then,

$$\forall C \in \mathcal{C} : \exists L \in \mathcal{L} : C \subseteq S(L)$$

In plain words, while the senses of a homonym may be divided between multiple clusters, no cluster should contain senses from different homonyms.

## 3 Homonym Data

In order to provide experimental evidence for our homonym hypotheses, we need a large set of "gold" homonyms, as well as a mapping between those homonyms and the sense annotations in existing corpora. Since no such resource is publicly available, we create our own collection of English homonyms (see Table 2). In this section, we present a binary typology of homonyms, our methodology for creating a list of homonyms, and the method for mapping those homonyms onto the WordNet sense inventory.
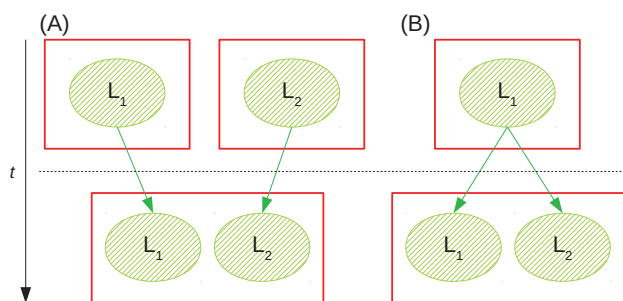
Figure 2: A schematic illustration of the diachronic distinction between two types of homonyms. Circles represent lexemes; boxes represent words.

## 3.1 Typology of Homonyms

There are generally two ways of defining homonyms. In linguistics (and in this paper), homonyms are considered to be distinct lexemes that happen to share the same form (Murphy and Koskela 2010). In lexicography, homonymy is sometimes defined more narrowly, by additionally requiring the etymological origins of the lexemes to be different (Stevenson 2010). Homonyms can therefore be divided into two types: those that satisfy the requirement of different origins, and those that do not. Due to the lack of commonly-accepted terminology, we refer to these two types of homonyms simply as Type-A and Type-B, respectively.

The two types of homonyms, which are schematically illustrated in Figure 2, stem from different diachronic phenomena. Type-A homonyms arise from a convergence of distinct words into a single form. This can occur through the process of sound change or inter-lingual borrowing. For example, both the Old English word *cæg* "locking implement" and the 17th-century Spanish borrowing *cayo* "island" evolved into the modern English *key*. Type-B homonyms, on the other hand, arise when a single lexeme splits into two lexemes due to the process of semantic drift. For example, the two meanings of *staff*, "pole" and "people", have developed from a single etymon, which is attested in Old English as *stæf*. Importantly, as native speakers are generally unaware of the etymological history of words, these two types of homonyms are indistinguishable in the synchronic analysis of languages (Lyons 1995).

The crucial methodological advantage of Type-A homonyms is that they can be objectively identified by consulting existing etymological dictionaries. Even though the process of compiling an exhaustive list of Type-A homonyms for any language is time-consuming, it is still much easier and less controversial than conducting psychological experiments with human subjects (Brown 2008), or obtaining consensus within teams of linguistic experts (Weischedel et al. 2013). We have accomplished this task for English by creating a homonym resource that we describe next.

| POS | Origin | Gloss | French |
|-----|--------|-------|--------|
| N,V | Old French *espan* | distance | *portée* |
| N,V | Low German *spannen* | rope | *filin* |
| Adj | Old Norse *spán-nýr* | clean | *impeccable* |
| V | Old English *spinnan* | rotate | *tourné* |

Table 2: Sample entries of the homonym resource, which correspond to six homonyms of the English lemma *span*.

## 3.2 List of Type-A Homonyms

The new homonym resource[1], which enables us to empirically test our homonym hypotheses, contains words that represent multiple lexemes with distinct etymological origins. We compiled the list by collecting all homonyms that we could find in dictionaries, including the English Oxford Living Dictionary[2] and the Concise Oxford Dictionary of English Etymology[3]. We include all homonyms that at some point during language evolution existed as separate words, even those that can be traced to a single proto-word. For example, we include the homonyms of the noun *sole* ("undersurface" vs. "fish") because of their distinct histories, even though both ultimately come from Latin *solea* "sandal".

Table 2 shows sample entries from our resource. The list contains 2759 Type-A homonyms that correpond to 804 lemmas, 1601 unique lemma/POS pairs, and 1967 distinct etymologies. The number of distinct etymologies per lemma ranges from two to six. Each entry includes etymological information (the form and the language of origin), and a list of possible parts of speech (noun, verb, adjective, adverb). For the purpose of disambiguation in subsequent stages of annotation, each entry was manually assigned a brief English gloss, as well as a single French translation. We excluded from our list all proper nouns and abbreviations.

About two dozen of the homonymous words in our resource represent homographs, which are homonyms that differ in pronunciation. For example, the noun 'bass' is pronounced [bæs] or [bes] depending on whether it refers to a fish or a musical instrument, respectively. Although most of the dictionary words with alternative pronunciations appear to involve Type-A homonyms, we found a number of exceptions. They include Type-B homonyms (e.g. 'pension'), polysemous words (e.g. 'undertaking'), common vs. proper nouns (e.g. 'job'), matching word-forms of distinct lemmas (e.g. 'putter'), as well as pronunciation variants (e.g. 'puissance'). Since our focus is on written language, our resource excludes homophones, such as 'cellar' vs. 'seller'.

Although we make no claim about the completeness of our homonym resource, we consider it to be representative of English homonyms in general. This is based on the fact that Type-A and Type-B homonyms cannot be distinguished without access to etymological expertise.

---

[1] *https://webdocs.cs.ualberta.ca/~kondrak*
[2] *https://en.oxforddictionaries.com*
[3] *http://www.oxfordreference.com*

### 3.3 Mapping WordNet Senses to Homonyms

In order to test our homonym hypotheses, we must be able to convert the existing word sense annotations into homonym annotations. For example, we need to know which homonym from our list is represented by a word token *spans* which is sense-annotated as "two items of the same kind" in some corpus. The standard sense inventory for WSD is WordNet. In this section, we describe our method of mapping the homonyms in our new resource to WordNet senses.

Because of the large number of fine-grained senses in WordNet, it was not practical to directly map each WordNet sense of each homonymous word to the corresponding homonym. Instead, we made use of the existing clustering (Navigli 2006), which was created by automatically mapping WordNet 2.1 senses to more coarse-grained senses defined by the Oxford Dictionary of English (ODE). Our 804 homonymous lemmas correspond to 2644 sense clusters, which contain 5361 senses. We manually mapped each cluster of senses to a single homonym on the basis of their WordNet sense glosses.

The resulting mapping is imperfect for two reasons. First, the ODE clustering itself is not always correct, which sometimes results in homonymous senses being placed in the same cluster. Second, our human annotator made some errors in mapping clusters to homonyms. We performed the following validation experiment in order to estimate the accuracy of the overall mapping. A second annotator performed a direct mapping of 268 WordNet senses corresponding to a random sample of 77 homonymous words, without any reference to the ODE clustering. We found that the two independent mappings of the 268 senses differed in only 17 instances, which implies that the overall error rate has an upper bound of 6%.

The errors in the sense-to-homonym mapping are a source of "false alarms" in the experiments described in Section 4. We are confident in our ability to determine which of the apparent exceptions are actual exceptions to our hypotheses by careful analysis of the available data. While the distinction between homonymy and polysemy can be highly subjective, the mapping of WordNet senses to known homonyms is much easier, as confirmed by our validation experiment described above.

## 4 Homonym Evidence

In this section, we describe the experiments that test the four hypotheses formulated in Section 2 using the full set of homonyms in our new homonym resource from Section 3.

### 4.1 SemCor and Translations

For testing the OHPD and OHPC hypotheses, we use Sem-Cor (Miller et al. 1993), a large sense-annotated English corpus which was created as part of the WordNet project (Petrolito and Bond 2014). In particular, we adapt the version of SemCor from (Raganato, Camacho-Collados, and Navigli 2017).[4] The number of word tokens, types, and senses are in Table 3 (words are defined as lemma/POS pairs)

|  | SemCor | MSC | JSC |
|---|---|---|---|
| Word tokens | 226,034 | 92,992 | 58,257 |
| Word types | 20,399 | 11,451 | 8,445 |
| WordNet senses | 33,308 | 17,875 | 12,516 |

Table 3: The size of the English side of each corpus.

For testing the OHPT hypothesis, we require not only sense annotations, but also the corresponding translations. At the minimum, we need a large word-aligned bitext that has both sense and part-of-speech annotations on the source side, and lemma annotations on both sides. In addition, the sense inventory has to be the same as the one in our homonym resource. Although such resources are rare, we managed to adapt two bitexts to meet these requirements: MultiSemCor (Bentivogli and Pianta 2005), and JSemCor (Bond et al. 2012). These corpora, which we refer to as MSC and JSC, contain partial word-aligned translations of SemCor into Italian and Japanese, respectively.

### 4.2 WordNet

The use of WordNet presents a number of technical challenges. For the purpose of replicability, we describe here two major issues.

The first issue concerns two distinct conventions for referring to individual WordNet senses: *sense keys* (used in Sem-Cor, JSC, and the ODE clustering) and *sense numbers* (used in MSC and OntoNotes). We converted the former into the latter using the WordNet::SenseKey package.[5] Because the mapping is not always one-to-one, 16 out of 60,655 WordNet senses in the ODE clustering had to be excluded; however, none of the affected words are in our homonym resource.

The second issue is the mapping between different WordNet versions. We converted the sense keys from WordNet 2.1 – the version of WordNet used in the clustering described in (Navigli 2006) – to WordNet 3.0 – the version used by all other resources in this paper – using WordNetMapper.[6] The package failed to map 551 out of 60,655 senses in the ODE clustering, which resulted in 22 WordNet senses being excluded from our homonym resource. Due to these issues, we decided not to further map all WordNet senses in our resources to WordNet 3.1.

### 4.3 One Homonym per Translation

The OHPT hypothesis characterizes the relationship between homonymous words and their translations in another language. We validate the hypothesis on two language pairs using the annotated bitexts described in Section 4.1.

In the experimental evaluation, we compute the percentage of type-level instances that are consistent with the OHPT hypothesis. For each English word (i.e. lemma/POS pair) that appears in our homonym resource, we identify the set of its translations on the target side of the bitext. Each unique

---

[4] *http://lcl.uniroma1.it/wsdeval*

[5] *https://metacpan.org/release/LINAS/WordNet-SenseKey-1.03*
[6] *https://github.com/cltl/WordNetMapper*

| # | Hypothesis | Focus | Corpus | Instances | Exceptions apparent | actual | Support (in %) |
|---|---|---|---|---|---|---|---|
| 1 | OHPT | translations | MSC (Italian) | 1093 | 7 | 1 | 99.9 |
| 2 | OHPT | translations | JSC (Japanese) | 1093 | 3 | 2 | 99.8 |
| 3 | OHPD | documents | SemCor | 2126 | 14 | 9 | 99.6 |
| 4 | OHPC | collocations | SemCor | 522 | 16 | 11 | 97.9[7] |
| 5 | OHPSC | sense clusters | OntoNotes | 1578 | 23 | 2 | 99.9 |

Table 4: Summary of the evidence for the homonym hypotheses from our five experiments.

word/translation pair constitutes a single instance. An instance is consistent with the OHPT hypothesis if and only if all of its occurrences in the bitext represent the same homonym. For example, the Italian translation 'gioco' corresponds to three different senses of the noun 'gioco' in MSC, but since all of them belong to the same homonym, this instance is consistent with OHPT.

The results of the evaluation on the MSC and JSC bitexts are shown in Rows 1 and 2 of Table 4. Coincidentally, MSC and JSC have the same number of unique word/translation pairs (1093). The two corpora contain only 3 actual exceptions to OHPT. The single actual exception in MSC involves the homonyms represented by the noun 'band' which is often translated in Italian as 'banda'. In this case, the homonymy in English ("ring" vs. "group") is mirrored by an analogous case of homonymy in Italian. The two actual exceptions in JSC involve the English lexical loans 'case' and 'club', which have the same Katakana written form regardless of the homonym they represent. We attribute these exceptions to the phenomenon of *parallel homonymy*, which may arise in the process of lexical borrowing.

In addition to the 3 actual exceptions, the experiment identified 7 exceptions that are caused by data errors in the two corpora. The data errors can be divided into four categories: (1) incorrect sense annotations in SemCor, e.g. *"the case of Jupiter"* annotated with the sense of "container"; (2) an incorrect sense translation in MSC: *flag* in the sense of "flower" translated as *bandiera* instead of *iride*; (3) errors in the ODE clustering, e.g. two homonymous senses of 'club' ("team" and "playing card") in the same cluster; (4) an error in our manual mapping between the ODE clustering and the homonyms: 'light' in the sense of "free from troubles" being mapped to the homonym "not dark". We conclude that the OHPT hypothesis is supported in over 99.8% of instances in either bitext.

In order to verify that partitioning of translations is a property of homonyms, and not simply of any sense clusters, we perform an additional experiment on MSC. We randomly select two sets of 20 words (i.e. lemma/POS pairs) from our homonym resource and the OntoNotes clusters, respectively. We consider only words that are represented in MSC by senses from exactly two homonyms or two OntoNotes sense clusters. None of the OntoNotes words occur in our homonym resource. This yields 40 words with a similar number of sense-annotated tokens: 6.80 per homonym, and 7.25 per OntoNotes cluster, on average. We find that 16 of the 20 homonym pairs, and 6 of the 20 OntoNotes cluster pairs exhibit strict translation partitioning in MSC. In to-

tal, there are 4 instances of overlapping translations between 4 homonym pairs (a subset of the 7 apparent exceptions in Table 4), and 17 such instances between 14 OntoNotes cluster pairs (3 cluster pairs share multiple translations). This result is statistically significant ($p < 0.005$) according to the $\chi^2$ test. We conclude that homonyms are significantly more likely to exhibit translation partitioning than OntoNotes sense clusters.

## 4.4 One Homonym per Discourse

The OHPD hypothesis predicts that all tokens of a given homonymous word in a discourse correspond to the same homonym. We validate the hypothesis on English SemCor (Section 4.1), taking each of its documents as a single discourse.

In the experimental evaluation, we compute the percentage of type-level instances that are consistent with the OHPD hypothesis. For each English word (i.e. lemma/POS pair) that appears in our homonym resource, we identify all its occurrences in the corpus. Each unique word/document pair constitutes a single instance. An instance is consistent with the OHPD hypothesis if and only if all of the occurrences of the word in the document represent the same homonym.

When a homonymous word occurs only once in a document, there is of course no possibility of an actual OHPD violation. However, we consider those instances to support the hypothesis as well, because the writer may have chosen to replace a homonym with one of its synonyms in order to avoid potential ambiguity.

The results of the evaluation are shown in Row 3 of Table 4. SemCor is divided into 352 documents, with an average of 642 sense-annotated open-class words per document. A careful analysis of the 14 apparent exceptions reveals that four of them are caused by sense annotation errors in SemCor (e.g., *sharp bow of a skiff* is annotated as "weapon for shooting arrows"), and one results from an error in the ODE clustering. The 9 actual exceptions involve the homonymous nouns 'bank', 'lead', 'list', 'port', 'rest', and 'yard', as well as the verb 'lie'. We conclude that fewer than 0.5% of instances in SemCor contradict the OHPD hypothesis.

## 4.5 One Homonym per Collocation

The OHPC hypothesis predicts that only one homonym of a word appears in any given collocation. Due to the broad

---

[7]This number is a lower bound estimate.

definition, wide variety, and large number of possible collocations, it is difficult to definitively establish the extent to which the OHPC hypothesis holds for a given corpus. Instead, we follow the methodology of (Yarowsky 1993) and (Martinez and Agirre 2000), who test the OSPC hypothesis by analyzing the performance of a supervised WSD system in which each feature corresponds to a distinct type of a collocation. The rationale is that the accuracy of the WSD system indicates the level of support for the hypothesis in the training corpus.

For the experimental evaluation, we adopt the IMS system of (Zhong and Ng 2010). IMS learns a separate classification model for each ambiguous word in the training data, with each class corresponding to one sense of the word. The system employs three types of features, which broadly correspond to different kinds of collocations: (1) the presence of specific content words in specific positions relative to the focus word; (2) the set of POS tags in the context of the focus word; (3) the presence of specific content words in the *bag-of-words* context of the focus word. We train IMS on English SemCor, and test on the concatenation of five benchmark datasets of (Raganato, Camacho-Collados, and Navigli 2017).

The results of the experiment strongly support the OHPC hypothesis. The test set contains 528 occurrences of words from our homonym resource. Six of those words, each appearing in one instance, are not attested at all in SemCor. IMS selects a sense of the correct homonym in 506 out of the remaining 522 instances. Of the 16 classification mistakes, three are attributable to errors in the ODE clustering, and two are due to the WordNet mapping issues described in Section 4.2. Thus, the effective accuracy of IMS on the homonymous words in the test set is 97.9%.

Analysis of the remaining 11 errors made by IMS shows that their principal cause is insufficient training data. For example, the noun 'match' in the sense of "piece of wood" occurs only once in the entire SemCor corpus, which prevents IMS from reliably recognizing this sense. Other obvious mistakes, such as *"follow the lead"* misclassified as "metal," are explained by the lack of training examples involving the collocations that occur in the test set. We conclude that the IMS accuracy on the test set should be interpreted as a lower bound for the applicability of OHPC.

### 4.6 One Homonym per Sense Cluster

We test our fourth hypothesis, OHPSC, by searching an existing resource for clusters that contain senses from distinct homonyms. We cannot perform this experiment on the ODE clustering because we use it to derive our mapping from WordNet senses to homonyms (Section 3.3). Instead, we run it on the high-quality, hand-crafted OntoNotes clustering[8], which previously used as a gold-standard by (Snow et al. 2007). The clustering includes 439 of the 1601 lemma/POS pairs that are listed in our homonym resource. Those words involve 2467 WordNet senses that are grouped into 1578 clusters, of which 1555 (98.5%) are found to contain no homonymous senses, as our hypothesis predicts.

---

[8]*https://catalog.ldc.upenn.edu/LDC2013T19*

We manually analyze the 23 clusters that appear to combine senses from distinct homonyms. The vast majority (21) of these apparent exceptions are artifacts of errors in the ODE clustering. The errors are easy to spot by native speakers because senses within a single cluster clearly correspond to distinct coarse-grained senses in ODE. In the remaining two cases, OntoNotes clusters two pairs of homonymous senses: (1) the noun 'tap' as "the sound made by a gentle blow" and "a faucet for drawing water," and (2) the verb 'pose' as "introduce" and "be a mystery to." Even though we find these two clustering decisions somewhat debatable, we treat them as actual exceptions to our hypothesis. We conclude that the OHPSC hypothesis is corroborated in over 99.8% of the OntoNotes clusters.

## 5 Conclusion

We have investigated the concept of homonymy, formulating four hypotheses that follow a common pattern. Taken together, our hypotheses suggest that, figuratively speaking, homonyms seem to repel each other, like particles with the same electric charge. The experiments performed using our new resource confirm that distinct homonyms are rarely observed in connection with a single translation, discourse, collocation, or sense cluster. In addition, they demonstrate that contraventions of the empirical predictions made by our theory more often than not identify errors in existing resources.

We envisage several directions for building upon the theoretical basis established in this paper. In order to extend our homonym resource, we plan to develop an operational method for identifying Type-B homonyms on the basis of translation sets involving multiple languages. We anticipate that translations extracted from parallel corpora will facilitate the creation of high-quality coarse-grained sense inventories via sense clustering. As a step towards this goal, we will investigate the problem of automated mapping between senses and translations.

## References

Apidianaki, M. 2008. Translation-oriented word sense induction based on parallel corpora. In *LREC*, 3269–3275.

Bansal, M.; DeNero, J.; and Lin, D. 2012. Unsupervised translation sense clustering. In *NAACL*, 773–782.

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

Bond, F.; Baldwin, T.; Fothergill, R.; and Uchimoto, K. 2012. Japanese SemCor: A sense-tagged corpus of Japanese. In *Global WordNet Conference*, 56–63.

Brown, S. W. 2008. Choosing sense distinctions for WSD: Psycholinguistic evidence. In *ACL*, 249–252.

Carpuat, M. 2009. One translation per discourse. In *Workshop on Semantic Evaluations*, 19–27.

Chan, Y. S.; Ng, H. T.; and Chiang, D. 2007. Word sense disambiguation improves statistical machine translation. In *ACL*, 33–40.

Dandala, B.; Hokamp, C.; Mihalcea, R.; and Bunescu, R. 2013. Sense clustering using Wikipedia. In *RANLP*, 164–171.

Diab, M., and Resnik, P. 2002. An unsupervised method for word sense tagging using parallel corpora. In *ACL*, 255–262.

Fellbaum, C. 1998. WordNet: An on-line lexical database and some of its applications. *MIT Press*.

Gale, W. A.; Church, K. W.; and Yarowsky, D. 1992. One sense per discourse. In *Workshop on Speech and Natural Language*, 233–237.

Hovy, E.; Marcus, M.; Palmer, M.; Ramshaw, L.; and Weischedel, R. 2006. OntoNotes: The 90% solution. In *NAACL*, 57–60.

Ide, N., and Wilks, Y. 2007. Making sense about sense. In Agirre, E., and Edmonds, P., eds., *Word Sense Disambiguation: Algorithms and Applications*. Springer. 47–73.

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

Katamba, F. 1993. *Morphology*. Macmillan Press.

Krovetz, R. 1998. More than one sense per discourse. *NEC Princeton NJ Labs., Research Memorandum* 23.

Liu, F.; Lu, H.; and Neubig, G. 2018. Handling homographs in neural machine translation. In *NAACL*, 1336–1345.

Lyons, J. 1995. *Linguistic semantics: An introduction*. Cambridge University Press.

Mancini, M.; Camacho-Collados, J.; Iacobacci, I.; and Navigli, R. 2017. Embedding words and senses together via joint knowledge-enhanced training. In *CoNLL*, 100–111.

Martinez, D., and Agirre, E. 2000. One sense per collocation and genre/topic variations. In *EMNLP*, 207–215.

McCarthy, D.; Apidianaki, M.; and Erk, K. 2016. Word sense clustering and clusterability. *American Journal of Computational Linguistics* 42(2):245–275.

Mel'čuk, I. 2013. *Semantics: From Meaning to Text*, volume 2. John Benjamins.

Miller, G. A.; Leacock, C.; Tengi, R. I.; and Bunker, R. T. 1993. A semantic concordance. In *ARPA Workshop on Human Language Technology*, 303–308.

Murphy, M. L., and Koskela, A. 2010. *Key Terms in Semantics*. London: Continuum.

Navigli, R. 2006. Meaningful clustering of senses helps boost word sense disambiguation performance. In *ACL*, 105–112.

Navigli, R. 2018. Natural language understanding: Instructions for (present and future) use. In *IJCAI*, 5697–5702.

Ng, H. T.; Wang, B.; and Chan, Y. S. 2003. Exploiting parallel texts for word sense disambiguation: An empirical study. In *ACL*, 455–462.

Petrolito, T., and Bond, F. 2014. A survey of WordNet annotated corpora. In Orav, H.; Fellbaum, C.; and Vossen, P., eds., *Global WordNet Conference*, 236–245.

Pilehvar, M. T.; Camacho-Collados, J.; Navigli, R.; and Collier, N. 2017. Towards a seamless integration of word senses into downstream NLP applications. In *ACL*, 1857–1869.

Raganato, A.; Camacho-Collados, J.; and Navigli, R. 2017. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *EACL*, 99–110.

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

Resnik, P., and Yarowsky, D. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural Language Engineering* 5(2):113–133.

Snow, R.; Prakash, S.; Jurafsky, D.; and Ng, A. Y. 2007. Learning to merge word senses. In *EMNLP-CoNLL*, 1005–1014.

Snyder, B., and Palmer, M. 2004. The English all-words task. In Mihalcea, R., and Edmonds, P., eds., *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 41–43.

Stevenson, A. 2010. *Oxford dictionary of English*. Oxford University Press, USA.

Taghipour, K., and Ng, H. T. 2015. One million sense-tagged instances for word sense disambiguation and induction. In *CoNLL*, 338–344.

Utt, J., and Padó, S. 2011. Ontology-based distinction between polysemy and homonymy. In *International Conference on Computational Semantics*, 265–274.

Weischedel, R.; Palmer, M.; Marcus, M.; Hovy, E.; Pradhan, S.; Ramshaw, L.; Xue, N.; Taylor, A.; Kaufman, J.; Franchini, M.; et al. 2013. OntoNotes release 5.0. *Linguistic Data Consortium, Philadelphia, PA*.

Yarowsky, D. 1993. One sense per collocation. In *Workshop on Human Language Technology*, 266–271.

Zhong, Z., and Ng, H. T. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *the ACL 2010 System Demonstrations*, 78–83.