UNIVERSITY OF MILANO - BICOCCA
DEPARTMENT OF COMPUTER SCIENCE, SYSTEMS AND COMMUNICATIONS
DOCTORAL SCHOOL IN COMPUTER SCIENCE - XXVIII CYCLE

Ph.D. Thesis

Cross-Lingual Mapping of Lexical Ontologies with Automatic Translation

Mamoun Abu Helou

Supervisors:	Dr. Matteo Palmonari
	Dr. Mustafa Jarrar
Tutor:	Prof. Stefania Bandini

2014/2015

UNIVERSITY OF MILANO - BICOCCA
DEPARTMENT OF COMPUTER SCIENCE, SYSTEMS AND COMMUNICATIONS
DOCTORAL SCHOOL IN COMPUTER SCIENCE - XXVIII CYCLE

Ph.D. Thesis

Cross-Lingual Mapping of Lexical Ontologies with Automatic Translation

Author: Mamoun Abu Helou

Supervisors: Dr. Matteo Palmonari
Dr. Mustafa Jarrar
Tutor: Prof. Stefania Bandini

Dissertation Committee: Prof. Stefania Bandini
Prof. Claudia Archetti
Prof. Alessandra Carbone

Thesis Reviewers: Prof. Domenico Beneventano
Dr. Catia Pesquita

2014/2015

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In the name of Allah, the Most Gracious, the Most Merciful

Abstract

In the Web, multilingual data are growing fast and exist in a large number of sources. *Ontologies* have been proposed for the ease of data exchange and integration across applications. When data sources using different ontologies have to be integrated, mappings between the concepts described in these ontologies have to be established. *Cross-lingual ontology mapping* is the task of establishing mappings between concepts lexicalized in different languages. Cross-lingual ontology mapping is currently considered an important challenge, which plays a fundamental role in establishing semantic relations between concepts lexicalized in different languages, in order to align two language-based resources; to create multilingual lexical resources with rich lexicalizations; or to support a bilingual data annotation.

Most of the cross-lingual mapping methods include a step in which the concepts' lexicalizations are automatically translated into different languages. One of the most frequently adopted approaches in the state-of-the-art to obtain automatic translations includes the use of *multilingual lexical resources*, such as machine translation tools, which have been recognized as the largest available resources for translations. However, translation quality achieved by machine translation is limited and affected by noise; one reason of this quality is due to the polysemous and synonymous nature of natural languages. The quality of the translations used by a mapping method has a major impact on its performance.

The main goal of this thesis is to provide an automatic cross-lingual mapping method that leverages lexical evidence obtained from automatic translations, in order to automatically support the decision in mapping concepts across different languages, or even to support semi-automatic mapping workflows. In particular, in establishing mappings between very large, lexically-rich resources, e.g., lexical ontologies. The major contributions of this thesis can be summarized as follows: I presents a classification-based interpretation for cross-lingual mappings; I analyze at a large-scale the effectiveness of automatic translations on cross-lingual mapping tasks; I classifies concepts in lexical ontologies based on different lexical characteristics; I proposes an automatic cross-lingual lexical mapping method based on a novel translation-based similarity measure and a local similarity optimization algorithm; finally, I implements a Web tool that supports a semi-automatic mapping approach based on the proposed method.

Acknowledgements

First, I'm so thankful to Almighty Allah, Most Gracious, Most Merciful, without his blessing this research would not have been done. I am most grateful for the unconditional support from my parents, wife, sisters, brothers, and children, this thesis is dedicated to them. I would like to thank my supervisors. My sincere thanks to Dr. Matteo Palmonari for his guidance, encouragement and support over the years. I would also like to express my gratitude to my tutor Prof. Stefania Bandini for providing me with the necessary support during this Ph.D. My greatest appreciation to all my wonderful friends who have shown much love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Chapter Overview

This chapter presents an overview of this thesis. In particular, the motivation of this research is discussed in Section 1.2. The research question addressed by this thesis is presented in Section 1.3. A list of objectives and goals derived from this research question are discussed in Section 1.4. The technical approach undertaken for this research is presented in Section 1.5, followed by a discussion of the contributions in Section 1.6. Finally, Section 1.7 presents an overview of the remaining chapters of this thesis.

1.2 Motivation & Thesis Scope

Over the last years, the Web has witnessed an enormous growth in the amount of multilingual data [48]. When data is published in different languages, this leads to increasing semantic heterogeneity and make semantic interoperability more challenging [16, 88, 46, 53]. Tim Berners-Lee has envisioned the Semantic Web as *"an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"* [13]. Since then, *ontologies*, as specifications of conceptualizations [50], have gained great attention in research as well as in industry for enabling knowledge representation and sharing. Ontologies are perceived as language-independent representations of concepts and their interrelations, thereby allowing intelligent agents and applications to access and interpret Web content automatically.

To meet the vision of the semantic web, one of the significant goals is enabling the access and integration of data published in the Web. When data sources that use different ontologies have to be integrated, mappings between the concepts described in the different ontologies have to be established. This task is also called *ontology mapping* [16, 37]. Automatic ontology mapping methods are introduced to ease this task by finding potential mappings and determining which ones should be included in a final alignment.

Research in the field of ontology mapping has largely been focused on dealing with ontologies that are lexicalized in the same natural language (i.e., mono-lingual ontology mapping). In the last few years, research has also focused on providing assistance and support in mapping scenarios where the ontologies involved are lexicalized in different natural languages. Given the limitations of existing mapping tools that focus on mostly mono-lingual mapping processes, there is a need for development of mapping techniques that can work with data sources (e.g., ontologies) in different natural languages. One way to enable semantic interoperability between ontologies in different natural languages is by means of cross-lingual ontology mapping [43].

Cross-lingual ontology mapping is the task of establishing mappings between concepts of a source ontology lexicalized in a language and concepts of a target ontology lexicalized in a different language [107]. Different ontology representation models have been proposed for the ease of data exchange and integration across applications. *Lexical ontologies* define the meaning of concepts by taking into account the words used to express them [52]: each concept is defined by one or more synonym *words* [79], which I refer to as *lexicalization of the concept*, and connected to other concepts by semantic relations. *Axiomatic ontologies* are represented in logic-based languages like OWL[1] and define concepts by means of logical axioms. They are usually defined by a set-theoretic semantics and support automated reasoning [55]. Several hybridizations of these two approaches have been also proposed [116].

Cross-lingual ontology mapping is currently considered an important challenge [46]. Cross-lingual mapping tasks can be used to establish semantic relations between

[1]http://www.w3.org/TR/owl-features/

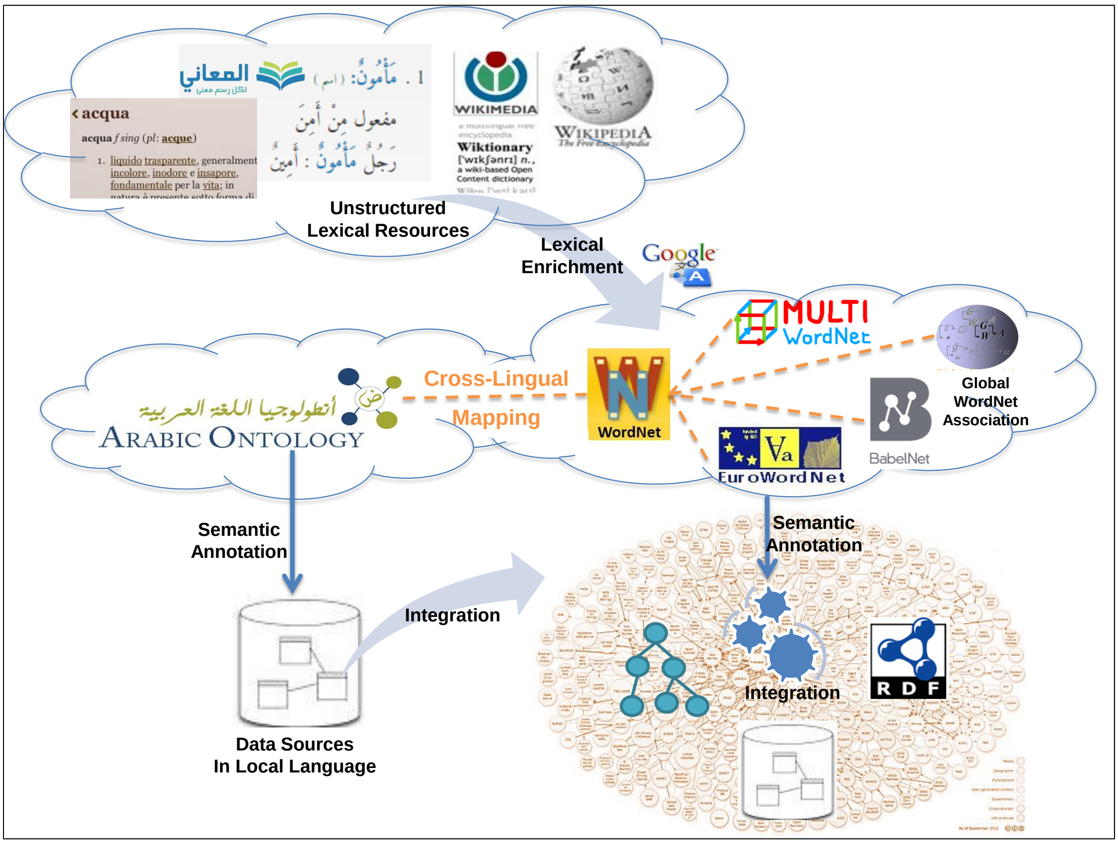


FIGURE 1.1: Cross-lingual mapping application scenarios

concepts lexicalized in different languages [111] in order to align two language-based resources [107, 44, 70, 111], to create multilingual lexical resources with rich lexicalizations [115, 93, 86], or to support a bilingual data annotation [105, 117]. Figure 1.1 illustrates few application scenarios in which cross-lingual mapping techniques can play a fundamental role to be achieved, which I explain further below.

The creation and enrichment of multilingual knowledge resources. Cross-lingual mapping methods can play a crucial role in bootstrapping the creation of large lexical ontologies (wordnets) and, for analogous reasons, in enriching existent ontologies [30, 86, 6]. The top part of Figure 1.1 demonstrates these processes. The most frequently adopted approaches to build and enrich multilingual knowledge resources are based either on mapping concepts lexicalized in different languages (merge model) or on translating the concepts' lexicalizations (expand model) [115]. The expand model was used more substantially than the merge model in approaches to automate the enrichment of multilingual wordnets and knowledge resources. One way to enrich an existing wordnet via the merge approach by mapping an unstructured or a weakly structured lexicon, e.g., a dictionary, to a structured reference ontology, e.g., the English WordNet [40]. For example, in the Arabic Ontology project [60, 6], the authors plan to use this approach to extend a core ontology manually created and mapped to the English WordNet. However, the mapping task incorporated in this approach is particularly challenging [1]: the lack of semantic relations between the concepts of an unstructured lexicon makes it difficult to disambiguate their meaning during the translation and the matching steps [103, 111]. An effective cross-lingual ontology mapping method can support the application of the merge model at large scale, thus supporting the construction and enrichment of multilingual knowledge resources. In addition, enriching the lexicalization of concepts in axiomatic ontologies with a set of synonyms is a well-established practice in ontology mapping [76, 103, 105, 39, 92, 28].

The semantic annotation. The Web constitutes of a large number of sources in different languages. For example, more than a million datasets have been published online as linked open data in 24 different languages in the linked open government data initiative (LOGD)[2]. In the *COMSODE*[3] project, several tables lexicalized in different languages have to be published on the Web after being annotated with ontologies and transformed in RDF. Data publishers would like to annotate their data with concepts lexicalized in their language as well as in English, in order to facilitate local citizens and integrate their data with the large amount of data published in English. A cross-lingual ontology mapping system may help them by facilitating bilingual data annotation [105, 117]. The bottom part of Figure 1.1 demonstrates this process.

Although cross-lingual ontology mappings approaches have been proposed in the last few years, cross-lingual mapping systems still perform significantly worse than mono-lingual mapping systems according to recent results in the Ontology Alignment Evaluation Initiative (OAEI) campaigns[4] [104], which suggest that cross-lingual ontology mapping is still a very challenging problem [111]. In what follows, I highlights main limitations for existing cross-lingual ontology mapping methods.

The semantic nature of mappings that cross-lingual ontology mapping methods are expected to find has not been sufficiently investigated [6, 43].

[2]http://logd.tw.rpi.edu/iogds_data_analytics , visited in Feb 2015.
[3]http://www.comsode.eu/
[4]http://oaei.ontologymatching.org/

Most of the cross-lingual ontology mapping methods include a step in which the concepts' lexicalizations of one ontology are automatically translated into the language of the other ontology [93, 115, 111]. The most frequently adopted approach in the state-of-the-art to obtain automatic translations is to use *multilingual lexical resources*, such as machine translation tools or bilingual dictionaries. However, a systematic and large-scale analysis of the effectiveness of translations in the context of cross-lingual mapping is missing [5]. In addition, in previous work, e.g., in the OAEI contest, the datasets used to evaluate cross-lingual mapping, i.e., the datasets in the multifarm track[5] [77], consist of alignments established between axiomatic ontologies of relatively small size and specific to the domain of conference organization(Cmt, Conference, ConfOf, Edas, Ekaw, Iasted, Sigkdd).

Methods proposed for cross-lingual ontology mapping [104], highly leverage the structural information available in the source and target ontologies. After a translation step, the similarity between the source and target concepts is computed by comparing their lexicalizations (e.g., string-based measures). These methods can not be directly adopted when structural information is not available.

Efficient matching methods have been introduced in mono-lingual ontology matching, e.g., [27], which computes the similarity among every source and target concept. However, this method is difficult to scale for very large mapping scenarios, e.g., mapping the Arabic Ontology to the English WordNet, which has about 120 thousand concepts.

Given the limitations of existing cross-lingual mapping methods, in this thesis I focus on cross-lingual mapping tasks when very large, lexically-rich language resources are considered. I will investigate the role of the translation evidence in supporting the decision of selecting the correct matches in an efficient way, in contrast to related work which use translations only as a step to transfer the problem into a mono-lingual mapping process [104]. Although the focus will be to leverage the lexicalizations of concepts, structural information should be smoothly incorporated when it is available.

[5] http://oaei.ontologymatching.org/2015/multifarm/index.html

1.3 Research Question

This research investigates to which extent evidence from automatic translations can be used to establish quality mappings between very large lexical resources.

Automatic cross-lingual matching methods can be used either to compute mappings automatically, even at the price of accuracy [31], or to support semi-automatic mapping workflows by recommending mappings to lexicographers [93]. The focus of this Ph.D. is to present an efficient cross-lingual lexical mapping method to map very large, lexically-rich resources, i.e., resources that associate each concept with a set of synonym words [79, 92]. In particular, to provide an efficient automatic mapping techniques, or even to support users, in establishing high quality mappings when contextual evidence (e.g., structural information) are limited [82, 117], or may not be even available, e.g., when an unstructured lexicon is matched against a structured ontology [60, 6].

To evaluate the approach I use cross-lingual mappings manually established (or validated) by lexicographers between four wordnets (Arabic, Italian, Slovene and Spanish) and the English WordNet. Using gold standards based on these wordnets has two main advantages. They contain a large number of mapped concepts, much larger, e.g., than the gold standards used to evaluate cross-lingual ontology mapping systems in the OAEI contest, and I take advantage of the lexical richness of concepts in theses resources to provide more in-depth analysis. The wordnets used in the experiments are also representative of different families of languages and of different ontology sizes, and not in a specific domain. To measure mappings quality, evaluation metrics such as precision, recall, F_1-measure are used from the state-of-the-art in ontology mapping field [103].

1.4 Objectives and Goals

To address the research question discussed in Section 1.3, the following objectives have been derived:

- Review of the state-of-the-art in cross-lingual ontology mapping and in other related tasks, such as enrichment of multilingual knowledge resources, cross-lingual word-sense disambiguation, and current approaches to the evaluate the performance of mapping methods.
- Investigate the semantic nature of cross-lingual mappings that cross-lingual ontology mapping methods are expected to find, i.e., define a theoretical interpretation of the meaning of these mappings.
- Study the effectiveness of automatic translations to support cross-lingual mapping tasks; analyze the impact of translations on retrieving a set of potential matches and the selection of the correct matches for inclusion in the final alignment.
- Provide an efficient cross-lingual mapping method that is specifically suited for establishing mappings between very large lexical resources.
- Evaluate the performance of the proposed mapping method.

1.5 Methodology & Technical Approach

At first, a state-of-the-art review (discussed in Chapter 3) is conducted in the field of cross-lingual ontology mapping and related tasks. I found that a formal interpretation (semantic) of cross-lingual mappings produced by the state-of-the-art mapping systems has not sufficiently investigated. First, I selected an appropriate interpretation of the mappings obtained as a result of the mono-lingual ontology mapping process. Then, I defined the semantics of mappings for cross-lingual ontology mapping, taking into account the lexicalization of concepts in the definition of the mapping (discussed in Chapter 4).

I also observed that, the quality of the translations used by a mapping method has a major impact on its performance. However, I found that a systematic and large-scale analysis of the effectiveness of automatic translations in the context of cross-lingual mapping is missing. To fill in this gap, I have analyzed at a large-scale the effectiveness of automatic translations to support cross-lingual ontology mapping tasks (discussed in Chapter 5). I studied four different large datasets, each consisting of a pair of mapped lexically-rich resources (wordnets), to cover four different families of languages. In order to have a comprehensive understanding of the translations impact on the mapping tasks, I defined a classification method for concepts based on different lexical characteristics.

The results of this study inspired the definition of a novel cross-lingual lexical ontology matching method (discussed in Chapter 6), where a lexical similarity measure based on evidence collected from translations is defined, and an efficient mapping selection technique is developed based on a local similarity optimization algorithm.

Since structural information might be, even partially, available in some mapping scenarios, I investigated the incorporation of these information along with the proposed method. I implemented a state-of-the-art structural matching method to evaluate the quality of mappings.

Automatic mapping methods are neither correct nor complete when compared to gold standards, thus the mapping process also requires users to validate the automatic mappings. I designed and implemented an interactive mapping application (discussed in Chapter 6), where users (lexicographers) can validate the generated mappings.

To evaluate the performance of proposed method state-of-the-art metrics are selected [103], in particular, measuring correctness and completeness of a set of mapping against a gold standard. Precision is used to evaluate the correctness and Recall is used to evaluate the completeness of a set of mappings. F_1-measure, which is the harmonic mean of precision and recall, is used to evaluate the overall quality of a set of mappings.

1.6 Contributions

The investigation of the outlined research question leads to the following contributions of this thesis, which also constitute the scientific accomplishment of the author. Five peer-reviewed scientific publications have been derived from this research, which are listed in Appendix A. The major scientific contributions of this Ph.D. can be summarized as follows.

Contribution I: *Cross-lingual mappings semantic.* The extension of concept is often used in many matching strategies [37, 90]. This approach is adapted and a classification-based semantics for cross-lingual ontology mapping is defined, taking into account the lexicalization of concepts. The classification task is viewed as disambiguation task, namely, the classification of a word as occurrence of a word sense in the sentence. Based on this lexical-based classification task an experimental setting is defined, through which the proposed mapping semantics can be evaluated and a gold standard dataset can be generated. Details on these activities are discussed in Chapter 4.

Contribution II: *Effectiveness of automatic translations on the cross-lingual mapping tasks.* I conducted a large-scale study on the effectiveness and quality of translations returned by translation resources to match concepts lexicalized in different languages. Four very large repositories of cross-lingual mappings are used, which include mappings from wordnets in four different languages (Arabic [97], Italian [93], Slovene [41], and Spanish [49]) to the English WordNet[40]. In this study concepts (synsets) are classified into different categories, based on different characteristics: word ambiguity (e.g., monosemous vs polysemous), number of synonyms (e.g., synonymful vs synonymless), and position in a concept hierarchy (e.g., leaves vs intermediate concepts). Using these classifications, the effectiveness of automatic translations is evaluated by studying the performance on the cross-lingual mapping tasks executed using automatic translations for different categories of concepts. First the coverage of translation resources and its impact on the candidate match retrieval task is analyzed. Next, the difficulty of the mapping selection task is studied using a baseline mapping selection method. Effectiveness of automatic translations is evaluated in terms of coverage and correctness, which are based on a comparison with translations considered perfect (correct) according to a gold standard. Moreover, details on this study are described in Chapter 5.

Contribution III: *Cross-lingual lexical matching method.* Inspired by the study (discussed in Chapter 5) I provide a purely lexical matching algorithm to support very large cross-lingual mapping between lexically-rich resources, where concepts can be identified by synsets, i.e., resources that associate each concept with a set of synonym words. The key idea of this algorithm is to use the results of word translations as evidence in order to map synsets lexicalized in different languages. I propose a *translation-based similarity measure* (TSM) inspired by a classification-based mapping semantics (discussed in Chapter 4). Further, I define a novel *local similarity optimization algorithm* (LSOA) to select the best matches for each source synset. To evaluate this approach I use wordnets in four different languages, which have been manually mapped to the English WordNet. Results show that despite the proposed method uses only lexical evidence about the concepts, it obtains good performance and significantly outperforms two baseline methods. I also combined this approach with a structural matching method. The results are improved and the structural matching method is smoothly incorporated in the similarity evaluation step of the proposed method. The corresponding details are discussed in Chapter 6.

Contribution IV: *Interactive mapping application.* Mappings found by a fully automatic methods are neither correct nor complete w.r.t gold standards. For such cases, I present an interactive mapping Web tool called ICLM (Interactive Cross-lingual Mapping), which aims to improve an alignment through incorporating multiple users in the validation process. ICLM tries to reduce users effort in validating the mapping tasks; It distributes the mappings on some users based on the difficulty of the tasks, which is estimated based on the lexical characterization of concepts under evaluation; and on how confident the automatic mapping method is. Accordingly, ICLM estimates the effort (number of users) needed to validate the mapping. The corresponding details are discussed in Chapter 7.

1.7 Thesis Overview

The remainder of this thesis is organized as follows.

- Chapter 2 introduces some core concepts and notation used throughout the rest of the thesis, which cover ontology, concept lexicalizations, cross-lingual mapping and translation tasks.
- Chapter 3 positions this thesis with respect to related works. It gives an overview of the state-of-the-art techniques in cross-lingual mapping, and presents a background knowledge on ontology mapping problem, translation resources, and mappings semantics.

- Chapter 4 discusses the semantics of cross-lingual mappings. It presents a classification-based semantic for cross-lingual ontology matching. An experimental setting is also introduced, through which the proposed mapping semantics can be evaluated and a gold standard dataset can be generated.

- Chapter 5 describes a large-scale study on the effectiveness of automatic translation resources on cross-lingual matching. It discusses the lexical characteristics of concepts, and introduces the evaluation measures used to evaluate the impact of the translation resources on the candidate match retrieval and mapping selection tasks.

- Chapter 6 discusses a cross-lingual lexical ontology matching method. It describes a novel lexical-based cross-lingual similarity measure, and a purely lexical and efficient mapping selection algorithm with local similarity optimization. A large-scale experiment is provided in order to quantitatively investigate the performance of the approach with respect to alternative methods.

- Chapter 7 presents a semi-automatic mapping approach that supports the validation process provided by multiple lexicographers. An experimental setting is also introduced, through which the effectiveness and usability of the approach is evaluated.

- Chapter 8 finally concludes this thesis with a summary of the research objectives achieved and contributions of this research, and outlines the potential future work and impact of the research topic.

Chapter 2

Foundations and Technical Background

2.1 Chapter Overview

This chapter introduces the foundations and the technical background for the work presented in this thesis. First, in Section 2.2, I define what is an ontology. Further, in Section 2.3 I explain the notion of *concept* that I adopt throughout this thesis, and I describe the lexicalization of concepts. In Section 2.4, I define the translation tasks and describe lexical resources that are used as sources of translations in this thesis. Finally, in Section 2.5 I define the cross-lingual mapping tasks and describe cross-lingual mappings datasets, which I used as gold standards throughout this thesis.

2.2 Ontologies

Ontologies have gained a lot of attention in recent years as tools for knowledge representation [103, 90]. Ontologies can be defined as a structured knowledge representation system composed by: *classes* (or concepts or topics), *instances* (which are individuals which belong to a class), *relations* (which link classes and instances, thus, allowing to insert information regarding the world to be represented in the ontology), and *terms* (which are the lexical representation, or labels, of the ontology elements in a given natural languages) [16].

As introduced in Chapter 1, matching systems consider different type of ontologies to be matched [81], thus, I can specialize the definition in several ways to reflect different type of ontologies. For instance, the different way an ontology is often used in ontology matching problem encompasses both lexical and axiomatic ontologies, described below.

In this thesis, I use the notion *ontology* to convey all the mentioned types, otherwise, I specifically differentiate the concerned type. Next, I explain the notion of concept in lexical and axiomatic ontologies, which I adopt throughout this thesis.

2.3 Concepts in Lexical and Axiomatic Ontologies

Concepts are the constituents of thoughts [72]. The relation between natural language and thought is much debated. For example, some maintain that concepts are independent from the language [42, 94] while other believe that concepts require natural language to exist [20, 106]. However, natural language plays a major role in expressing concepts in many computational knowledge representation systems proposed to support natural language processing, information retrieval and data integration tasks. *Ontologies* are among these computational knowledge representation systems. Two different kinds of ontologies can be distinguished.

In *lexical ontologies*, the meaning of concepts is primarily defined in relation to the words that can be used to express them. For example, in order to represent the concept "table", with reference to the object used to eat a meal, the set of words used to refer to this concept are specified. Lexical ontologies include domain thesauri, and wordnets, the most popular of which is the English WordNet [79, 40]. In *axiomatic ontologies* (or, logical ontologies) the meaning of concepts is defined by axioms specified in a logical language, e.g., First Order Logic, which are interpreted as constraints over mathematical structures and support automated reasoning [55]. Examples of logical ontologies include web ontologies defined in RDFS[1] or OWL, but an annotated database schema or a spreadsheet can also be considered an ontology based on this broad definition [118, 95, 82, 117]. For example, to represent the afore-mentioned concept "table", I can define it as "a

[1]http://www.w3.org/TR/rdf-schema

piece of furniture having a smooth flat top that is usually supported by one or more vertical legs" in a logical language. The intended interpretation of this concept can be every such table that had ever existed in the world, or, more specifically, a list of products of type "table" described in a spreadsheet [82, 117].

Many hybrid approaches also exist. For example, efforts to assure certain logical properties of relations represented in lexical ontologies can be found in KYOTO [116]. YAGO is a logical ontology that integrates many concepts from the English WordNet [109]. WordNet concepts used to annotate a database schema can be given a formal interpretation and used to support database integration [105].

As a matter of fact, despite several differences, concepts modelled in lexical, axiomatic, or hybrid ontologies *share two important features*. First, concepts are organized in *subconcept graphs*, i.e., hierarchies, partially ordered sets, or lattices that define the relations between concepts based on their generality. These relations are referred to as subconcept relations in axiomatic ontologies, while different relations can be represented in lexical ontologies, e.g., hyponymy/hypernymy. Second, in every ontology concepts have lexical descriptions that may include a set of synonym words. Of course, while synonyms are first class citizens in lexical ontologies and are available for a large number of concepts, their availability is more limited in axiomatic ontologies. However, a step to enrich the concept lexicalizations of logical ontologies with synonyms extracted from dictionaries and other lexical resources is introduced in many ontology mapping approaches so as to exploit lexical matching algorithms [103, 90, 105, 39].

Next, I discuss the lexical characterization of concepts that I adopt throughout this thesis.

2.3.1 Lexicalization of Concepts

In this thesis, I consider a general definition of ontologies, focusing on the lexical characterization of concepts, and on the relations between natural language words used in concepts. This choice is motivated by the observation that even ontology matching systems that look into the semantics of axiomatic ontologies, e.g., LogMap [63], use concept lexicalizations to retrieve candidate matches for concepts in a source ontology. For this reason, I borrow several definitions from lexical ontologies like WordNet [79] and use their terminology throughout the paper.

Slightly altering the terminology (but coherently with WordNet), **words** are lexemes associated with a concept. A word is called *simple* when it contains one token, e.g, "table", and is called *collection*[2] when it contains more tokens, e.g., "tabular array".

Wordnet organizes natural language words into *synonym* sets, called as **synsets**. Each synset represents one underlying *concept*, i.e., a set of words (synonyms) that share the same meaning in a given context. If W is the set of words represented in a wordnet, a synset $s \subseteq \mathcal{P}(W)$ is a set of words $s = \{w_1, ..., w_n\}$.

A synset can contain one word (*synonymless*) or many words (*synonymful*). I use "concept" and "synset" interchangeably in the rest of the thesis. Depending on the specific case, I use two notations for concepts: the set notation $\{w_1, ..., w_n\}$ which is used when I need to make explicit reference to the words contained in the synset, while a symbol notation s is used when this reference is not needed. I also use the set notation $w \in s$ to state that word w is contained in synset s. The set of words contained in the concept is also called its *lexicalization*. I use a superscript to specify the natural language used in concept lexicalizations when needed, i.e., w^L, s^L, or WN^L represent a word, a synset and a wordnet respectively lexicalized in the language L.

In addition to *lexical relations*, which link individual words (e.g., *synonymy*, *antonymy*), most of the wordnets support *semantic relations*, which link concepts. *Hypernymy* and *hyponymy* are the most important semantic relations in wordnets. They are defined one as the inverse of the other one and determine the subconcept graph in wordnets. For example, the synset {table, tabular array} is hyponym of the synset {array}, while the synset {array} is hypernym of the synset {table, tabular array}. WordNet includes brief definitions (glosses) and a short phrase illustrating the words' usage. Figure 2.1 shows an excerpt of the English WordNet[3].

A word is *polysemous*, i.e., has multiple meanings (or *senses*) when it is a member of many synsets. Throughout the thesis, I use the superscript "+" on the right-hand corner of a word, e.g., *"board$^+$"* to indicate a polysemous word. A word is

[2] An alternative name used instead of *collection* is *multiword expression* (MWE), which is frequently used. In particular, in the literature about machine translation tool evaluation [98], I use *collection* to be coherent with WordNet terminology that is used throughout the paper.

[3] Though this thesis I use WordNet version 3.0.

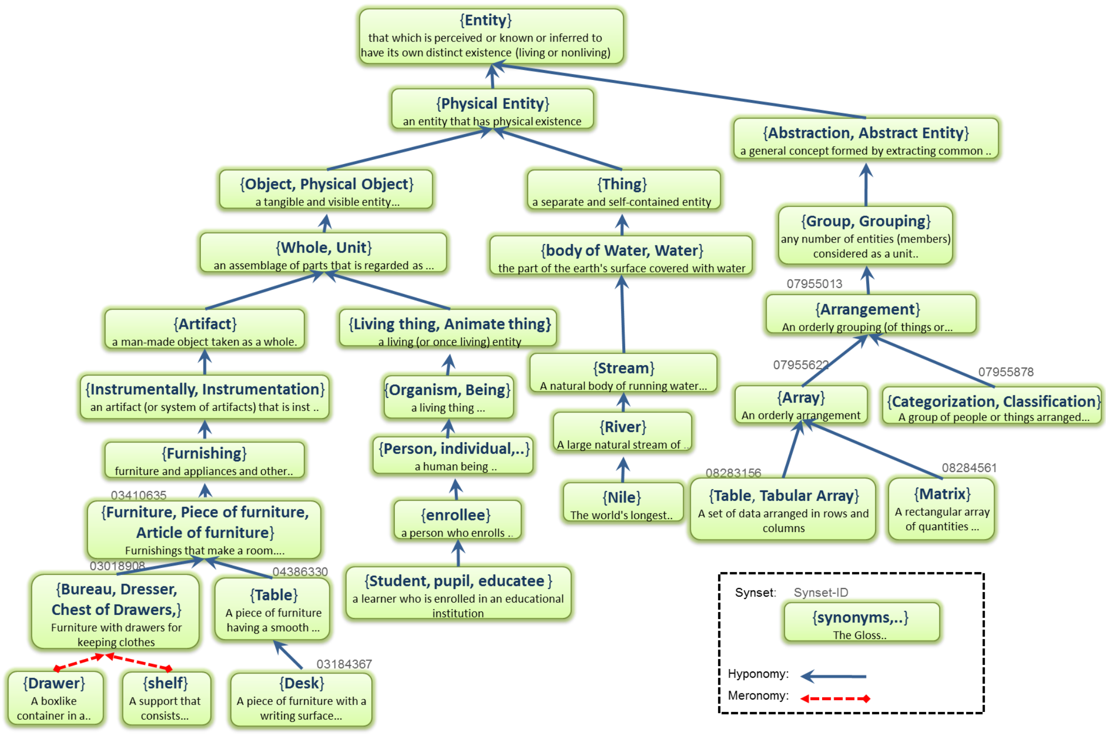


FIGURE 2.1: An excerpt of the English WordNet

monosemous, i.e. has only one meaning when it is a member of only one synset. For example, the English WordNet has eight senses for the word *"table$^+$"*; one of these senses means "a set of data arranged in rows and columns", which has the word *"tabular array"* as a synonym word. Another sense means "food or meals in general", which has the word *"board$^+$"* as a synonym word.

Given the set of words W and the set of synsets S defined in a wordnet WN, the function $senses : W \mapsto \mathcal{P}(S)$ returns the set of synsets that a word belongs to, defined by $senses(w) = \{s|w \in s\}$. I define the set of **word senses** in a wordnet as $WS = \{< w, s > |s \in senses(w)\}$, i.e., the set of couples $< w, s >$, such that s is a sense of w (in a given context). An important observation is that the number of word senses is higher than the number of synsets since all the associations between words and synsets are considered.

Example 2.1. The word "table$^+$" has eight senses in the English WordNet, $sense^{En}(table)$ = {{ table$^+$, tabular array },{table$^+$},{table$^+$},{mesa$^+$, table$^+$}, {table$^+$},{board$^+$, table$^+$}, {postpone, prorogue$^+$, hold over$^+$, put over$^+$, table$^+$, shelve$^+$, set back$^+$, defer$^+$, remit$^+$, put off$^+$ }, {table$^+$, tabularize, tabularise, tabulate$^+$}}[4].

2.4 Translation Tasks

Translating words of one language into words of another language is crucial in context of cross-lingual concept mapping [43], in particular, in the candidate match retrieval step. For the sake of clarity, I consider two translation tasks: translation of single words and translation of synsets. Translations are based on external resources, e.g., a machine translation tool or a dictionary built using a multilingual knowledge system.

Word-translation of a word w^{L_1} into a target language L_2 with a translation resource D is defined as a function $wTrans_D^{L_2} : W^{L_1} \mapsto \mathcal{P}(W^{L_2})$, which maps a word w^{L_1} into sets of words in a target language L_2.

Synset-translation of a synset s^{L_1} into a target language L_2 with a translation resource D is defined as a function $sTrans_D^{L_2} : S^{L_1} \mapsto \mathcal{P}(\mathcal{P}(W^{L_2}))$, which maps a synset s into sets of sets of words, each of which is the output of word-translation of some $w \in s$. The synset-translation function is defined as follows:

$$sTrans_D^{L_2}(s^{L_1}) = \{wTrans_D^{L_2}(w^{L_1}) \mid w \in s^{L_1}\} \tag{2.1}$$

Example 2.2. Figure 2.2 illustrates the synset-translation of the Italian synset $\{tavola^+, tabella\}^{It}$ into English can be given as follow: $sTrans_D^{En}(\{tavola^+, tabella\}^{It})$ $= \{wTrans_D^{En}\ (tavola^{+,It}),\ wTrans_D^{En}\ (tabella\ ^{It})\} = \{\{table,\ board,\ plank,\ panel,$ $diner,\ slab\},\ \{table,\ list\}\}$.

Observe that in the definition of synset-translation (eq 2.1) function, I do not make the set union of the outputs of every word-translation applied to the words in a synset. Instead, using Eq.2.1, the output of the synset-translation function can be written as *multiset union* of the sets returned by every word-translation. For instance, in Example 2.2, $sTrans_D^{En}\ (\{tavola^+,\ tabella\}^{It}) = \{table^{(2)},\ board^{(1)},$ $plank^{(1)},\ panel^{(1)},\ diner^{(1)},\ slab^{(1)}\}$, superscript numbers between brackets indicate the frequency count of the words in the translation set. Similarly, *"$table^{(2)}$"* means

[4] The senses definitions can be found on-line at http://wordnetweb.princeton.edu/perl/webwn?s=table

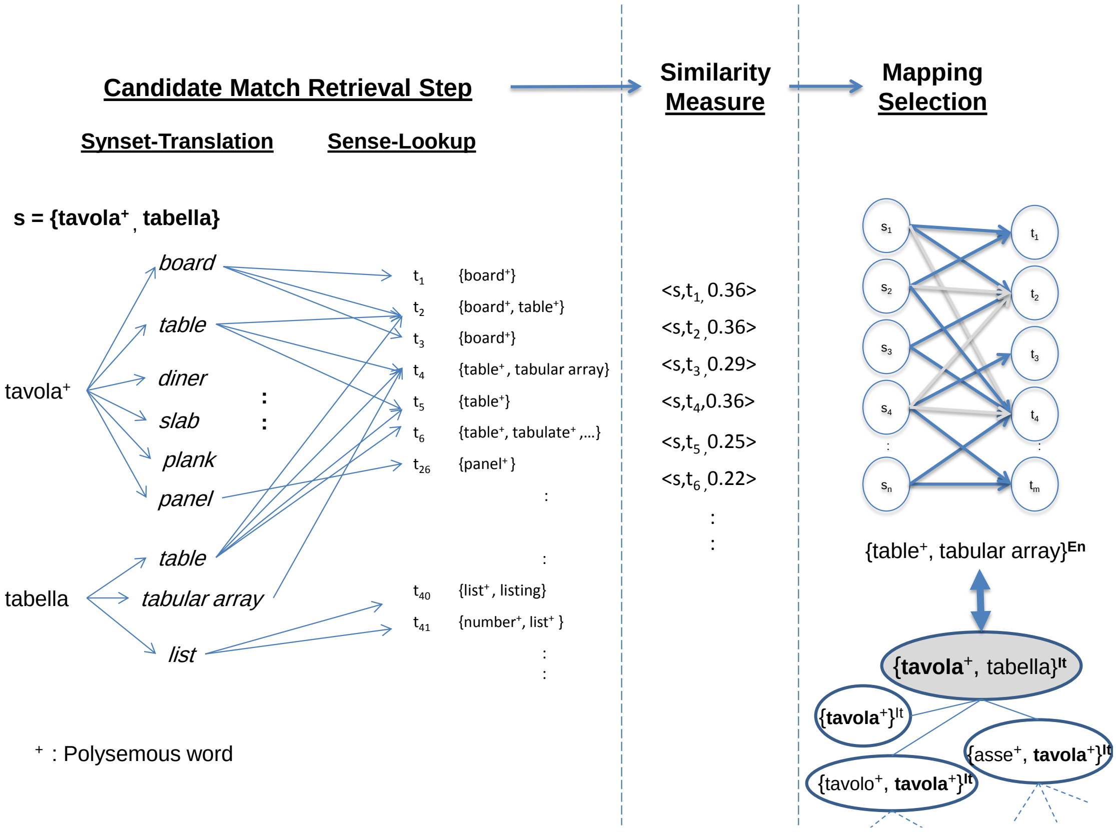


FIGURE 2.2: Cross-Lingual Mapping tasks

that the word "table" appears in two subsets, i.e., the word "table" is resulted as a translation of two synonym words in the source synset, which are "tavola" and "tabella". In this way I can count the number of word-translations that produce one word in the target language for a given synset-translation. These counts can be helpful to use the results of the synset-translation to perform the mapping selection step. For example, these counts can be used to weigh the candidate matches with a majority voting approach, similar to the one used in the experiments illustrated in Chapter 5, in Section 5.4.2, and in Chapter 6, in Section 6.5.

Next, I describe lexical resources used as sources of translations in this thesis; which I used to build bilingual dictionaries that are used in the candidate match retrieval tasks.

2.4.1 Multilingual Resource for Translation

Automatic translations can be obtained using different kinds of multilingual machine-readable lexical resources. The selection of these resources depends on the level of information they encode, for instance, the quality (accuracy) of translations they provide, the lexical domains they cover. These resources include: dictionaries, thesauri, wordnets, machine translation tools, and Web-based collaborative multilingual knowledge resources (resources in which lexical knowledge is manually and collaboratively generated, e.g., Wikipedia).

In this thesis two multilingual lexical resources are used as sources of translations: *Google Translate* and *BabelNet*. Google Translate is a statistical machine translation tool. Different machine translation systems exist that could be used; for instance, rule-based systems (e.g., Apertium[5]), and statistical-based systems (e.g., UPC [73]). I used Google Translate because previous work suggested that it performs better than other Web translation services in the context of concept mapping [7, 89], and has been adopted by several matching systems including the ones evaluated in the OAEI [104]. Moreover, Google Translate is a generic statistical machine translation, domain-independent system, and covers a very large number of languages, including the ones considered in study. A common evaluation measure of the machine translation quality is BLEU (Bilingual Evaluation Understudy) [91], which is based on the n-gram precision model. Thus, this measure does not fit the context of word-to-word translation, the case that I am considering. A comparison between different machine translation tools is out of scope of this study. For a rich and comprehensive comparison of different machine translation tools I refer to [25].

BabelNet is arguably the largest state-of-the-art multilingual knowledge resource. BabelNet has integrated several Web-based collaborative multilingual knowledge resources (see Section 3.5). In addition, it makes different translation strategies available, which I want to evaluate indirectly in the study: sense-tagged sentence translation, direct machine translation of monosemous words, and translations obtained from Wikipedia to Wordnet mappings.

I used Google Translate and BabelNet to construct bilingual dictionaries for every pairs of non-English and English languages considered in the gold standard (see Section 2.5.1).

Google Translate service is accessible through an API that can be used to translate *sentence-to-sentence* and *word-to-word* for many pairs of languages. Figure 2.3

[5]http://www.apertium.org

```
{"sentences":
[{"trans":"table","orig":"tavola","translit":"","src_translit":""}],
"dict":
[{"pos":"noun","terms":["table","board","panel","plank","slab","diner"],
    "entry":[
        {"word":"table","reverse_translation":["tabella","tavolo","tavola","tavolino","prospetto","mensa"],"synset_id":[55801],"score":0.44374731},
        {"word":"board","reverse_translation":["bordo","consiglio","pensione","commissione","tavola","comitato"],"synset_id":[34656,55801,92690],"score":0.0724397
        {"word":"panel","reverse_translation":["pannello","quadro","gruppo","riquadro","tavola","giuria"],"synset_id":[34656],"score":0.01889704},
        {"word":"plank","reverse_translation":["tavola","asse","tavolone","pancone"],"synset_id":[92690],"score":0.0040234118},
        {"word":"slab", "reverse_translation":["lastra","piastra","pezzo","tavola","tavoletta","tavolo anatomico"],"score":0.00025720813},
        {"word":"diner","reverse_translation":["commensale","trattoria","tavola","cliente"],"score":8.7510656e-05}
        ],
    "base_form":"tavola","pos_enum":1}
],
"src":"it","server_time":5}
```


Figure 2.3: Google Translate response for the Italian word "tavola" to English

shows Google's word-to-word translation response in JSON[6] format for the Italian word "tavola" to English. Google returns the preferable (common) translation in the `trans` item. A list of possible translations is also given in the `dict` item, which is part-of-speech (*PoS*) tagged. Each translation word in the `dict` item has a `reverse translation` set and a `score`. The `reverse translation` is a set of potential synonym words for the input word. The `score` estimates the translation usage (e.g., common, uncommon, or rare translations).

The translation directions (e.g., It-to-En, and En-to-It) of machine translation tools are said to have different performance when applied in cross-lingual information retrieval tasks [74]. To ensure the largest possible coverage I compiled three bilingual dictionaries using Google Translate. I take into account the translation direction. I also collect translations provided in the `reverse translation` sets. To the best of my knowledge available matching systems consider translations returned only in the `trans` item. For each pair of non-English and English languages considered in the gold standard I build the following bilingual dictionaries: MT_fromEn uses the translations collected from English to non-English words; MT_toEn uses the translations collected from non-English to English words; MT merges translations collected for the other two dictionaries to ensure the largest possible coverage (with Google Translate). Observe that MT_fromEn and MT_toEn are subsets of MT.

BabelNet is a large multilingual semantic network, which integrates several lexical resources (see Section 3.5). Nodes, called *BabelNet synsets*, represent concepts or named entities, which are lexicalized in several languages. For instance, the Italian lexicalizations in a node represent an Italian synset, which represents an equivalent synset to its corresponding English lexicalization, which is a synset in the English WordNet. The translation of a given word using BabelNet in a source language (e.g., It) into a target language (e.g., En) is given by every word in the target

[6] http://www.w3schools.com/json/

TABLE 2.1: Translation settings

Bilingual Dictionary	Description
MT_fromEn	translations from English to non-English words using Google Translate
MT_toEn	translations from non-English to English words using Google Translate
MT	the union of *MT_fromEn* and *MT_toEn*
BN	all translations encoded in BabelNet except translation from Open Multilingual WordNet
BN_{core}	BabelNet core synsets translations
$MT\&BN_{core}$	the union of *MT* and BN_{core}
MT&BN	the union of *MT* and *BN*

language, which localizes the same nodes that are lexicalized with the input word. For example, the Italian word "tavola" is lexicalization of 15 nodes[7] (14 concepts, and 1 named entity). These nodes provide 25 possible translations (lexicalization) in English : {board, correlation table, place setting, plank, setting, table, tablet, gang plank, wood plank, plate, table setting, stretcher bar, Panel, Panel cartoon, Oil on panel, ..., etc}. Each word in this lexicalization may derive from one or many of the different resources integrated in BabelNet.

To analyze the impact of the different lexical resources integrated in BabelNet, I extracted, for every pair of non-English and English languages used in experiments in Chapter 5 and 6, two bilingual dictionaries from the BabelNet synsets. A first dictionary is extracted from BabelNet *core* synsets (called BN_{core}), which contain multilingual lexicalizations built from: sense-tagged sentences, monosemous word translation using Google Translate (monosemous words heuristic), and Wikipedia inter-lingual links. A second dictionary is extracted from BabelNet synsets (called *BN*), all synsets in BabelNet, which contain multilingual lexicalizations built from: BN_{core}, and lexicalization obtained from WikiData, Wikitionary, OmegaWiki, and Wikipedia redirection links. Observe that BN_{core} is a subset of *BN*. I excluded only the Open Multilingual WordNet (OMWN) lexicalizations [14], which I adopt as gold standards in this thesis[8].

I also merged translations from BN_{core} and *MT* dictionaries (called $MT\&BN_{core}$), and translations from *BN* and *MT* dictionaries (called *MT&BN*). In this way I can compare and evaluate the impact of different Web-based linguistic resources, BabelNet core synsets, and the machine translation tools on the cross-lingual mapping tasks. The bilingual dictionaries I use in this study are summarized in Table 2.1.

[7]http://babelnet.org/search?word=tavola&lang=IT

[8]BabelNet specifies from which source each word of a lexicalization is obtained. Thus, I could remove those lexicalizations that have been obtained uniquely from the mapped wordnets used as gold standards (i.e., if word is obtained from the mapped wordnets but also from another source, I keep the word in the lexicalization).

2.5 Cross-Lingual Mapping Tasks

Cross-Lingual Mapping has been defined in ontology matching as the task of finding mappings between concepts of a *source ontology* lexicalized in a language L_1 and concepts of a *target ontology* lexicalized in a language L_2 [107, 111]. Mappings can represent different relations holding between source and target concepts. If a specific mapping relation R is considered, the output of a mapping task is a set of triples $\langle s, t, w \rangle$, also called an *alignment*; where s and t are concepts respectively of the source and target ontologies, and $w \in [0; 1]$ is a *weight* that denotes the degree of confidence that the association relation R (e.g., equivalence relation) between the couples $\langle s_i, t_j \rangle$ holds with a confidence weight w_{ij}. A **cross-lingual mapping task** with a mapping relation R is composed of three main steps (or, sub tasks), which are graphically illustrated in Figure 2.2:

- **Candidate Match Retrieval**: For each source concept s, a set of *candidate matches* $T = \{t_1, ..., t_n\}$ lexicalized in L_2 is retrieved. Figure 2.2 shows the association between the Italian synsets $\{tavola^+, tabella\}$ and a set of candidate synsets in English. Two subtasks are performed sequentially: *synset-translation* (eq 2.1) that is used first, followed by *sense-lookup* that is applied to the words returned by synset-translation; every target concept that contains at least one word translation is included in T. Target concepts are obtained from a lexical resource (e.g., wordnet or thesaurus) in language L_2.
- **Similarity Evaluation**: A confidence weight for each candidate match compared to the source concept is computed using similarity measures (e.g., lexical similarity measure, explained in detail in Chapter 6, in Section 6.3).
- **Mapping Selection**: Given a set of candidate matches $T = \{t_1, ..., t_n\}$ (lexicalized in L_2) for a source concept s (lexicalized in L_1), I select a set of concepts $T' \subseteq T$ such that for each $t \in T'$, $R(s, t)$ holds. When $R(s, t)$ holds for some $t \in T'$, I say that t is a *correct match* for s, and that $< s, t >$ is a *correct mapping*. Typically, the relation $R(s, t)$ with the highest confidence weight is selected.

Observe that the candidate match retrieval step define an *upper bound* for the mapping selection step: a correct mapping can be selected only if the target of the mapping was retrieved as a candidate match. In Chapter 5, I investigate this aspect, and analyze the impact of the translation resources on the cross-lingual mapping tasks. In addition, *mapping selection* is a form of *disambiguation task* (discussed in Chapter 6, in Section 6.4): the correct meaning of a concept (the lexicalization of the concept), in the target language has to be chosen among different possible meanings. A larger number of candidate matches and little evidence for preferring one candidate over another are likely to make the selection problem more difficult. Experiments conducted in Chapter 5 investigate these issues.

In this thesis, I consider only equivalence mappings, i.e., mappings that specify that a source and a target concepts have equivalent meaning. It is often assumed that the cardinality of equivalence mappings is 1:1, i.e., for each source concept there is at most one correct match. A more in-depth analysis of the semantics of cross-lingual mappings is discussed in Chapter 4.

Next, I describe cross-lingual mappings datasets, which I used as gold standards throughout this thesis.

2.5.1 Mapped Wordnets Used as Gold Standards

A *gold standard alignment* (or, **gold standard** for short), denoted by gs, is an alignment between synsets (concepts) in two lexical ontologies (e.g., wordnets) such that the mappings in the alignment are believed to be correct. In a gold standard alignment with cardinality 1:1, a synset in the source language have at most one *equivalent synset* in a target language. I use the predicate symbol "$\leftrightarrow$" to indicate that two synsets are equivalent (express the same meaning) in a gold standard. Using synset mappings in a gold standard gs, I define the *possible senses of a word w^{L_1} in a target language L_2*, denoted by $senses_{gs}^{L_2}(w^{L_1})$, as the senses in L_2 that are equivalent to the senses of w^{L_1} in its native language L_1:

$$senses_{gs}^{L_2}(w^{L_1}) = \{s^{L_2} \mid \exists s^{L_1}(w^{L_1} \in s^{L_1} \wedge s^{L_1} \leftrightarrow s^{L_2})\} \tag{2.2}$$

In this thesis, as gold standards, I use cross-lingual mappings manually established by lexicographers between four wordnets for Arabic [97], Italian [93], Slovene [41] and Spanish [49], and the English WordNet [79, 40]. These wordnets provide

Table 2.2: Size of the wordnets (gold standards)

	English	**Arabic**	**Italian**	**Slovene**	**Spanish**
Words	147306	13866	40178	39985	36880
Word senses	206941	23481	61588	70947	57989
Synsets	117659	10349	33731	42583	38702

high quality cross-lingual mappings and contain very large inventories of concepts. Their size in terms of words, word senses and synsets is reported in Table 2.2[9]. These wordnets have been built using different approaches and cover different families of languages: the *Germanic* languages (e.g., English), the *Romance* languages (e.g., Italian and Spanish), the *Slavic* languages (e.g., Slovene), and the *Semitic* languages (e.g., Arabic). Spanish, English, and Arabic are also among the top five spoken languages in the world[10], and their processing has gathered significant interest from the research community. Italian and Slovene represent two minority languages.

As said before, in the Introduction in Chapter 5, previous work in cross-lingual mapping have been mostly evaluated in the context of specific algorithms [44, 107, 104], with a limited number of gold standards, and for a limited number of languages. For instance, in OAEI contest, the multifarm dataset is composed of seven ontologies of the "Conference" domain (Cmt, Conference, ConfOf, Edas, Ekaw, Iasted, Sigkdd), which are of a small size (less than 150 concepts), and these ontologies have not been linguistically enriched (lexically-poor ontologies) [77]. Using gold standards based on these wordnets has two main advantages: they contain a large number of mapped concepts, much larger, e.g., than the ontologies used in the OAEI [104], and I can leverage the lexical characterization of concepts into different categories to provide a more in-depth analysis (discussed in Chapter 5). Theses wordnets are also representative of different families of languages and of different ontology sizes.

[9] The Arabic, Italian, and Slovene wordnets are obtained from http://compling.hss.ntu.edu.sg/omw, and the Spanish wordnet is obtained from http://adimen.si.ehu.es/web/MCR. All *lexical gaps* (synsets with no lexicalization) [115] are excluded.

[10] http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

Chapter 3

Background and State of the Art

3.1 Chapter Overview

The problem of finding mappings between concepts lexicalized in different languages has been addressed in the fields of cross-lingual ontology mapping and multilingual ontology construction. Cross-lingual word sense disambiguation is another research area where automatic translations have been proposed to solve matching problems. Related background and a state of the art review are presented in this chapter, with the goal of showing the role of concept lexicalizations and translation resources in cross-lingual concept mapping and related fields.

In this chapter, the ontology mapping problem is first introduced in Section 3.2. In Section 3.3 I describe the semantics of mappings in the monolingual ontology matching. I overview the classification-based approach for the weighted mappings proposed for monolingual ontology matching settings. In Section 3.4 I give an overview of the cross-language ontology matching in relation to the mono-language ontology matching definitions. Then, the construction of multilingual knowledge resources and cross-lingual word sense disambiguation are discussed respectively in Sections 3.5 and 3.6. Finally, in Section 3.7 I conclude this chapter with a summary. I discuss the contribution of this thesis to the evaluation of automatic translations for cross-lingual concept mapping and method proposed to process very large, lexically rich resources.

3.2 The Ontology Mapping Problem

With the rapid growth of the data on the Web, there is an increasing interest not only in sharing more data but also in sharing the semantics behind this data. The notion of a *Semantic Web* was proposed in [13] to deal with this massive growing amount of information and for machine understandable Web resources; to realize this, systems in the Web should be able to exchange information and services among other semantically. Thus, the semantics of one system should be exposed (in the Web) in such a way that other systems can understand it correctly and utilize it to achieve interoperability. Various ways have been introduced and proposed in order to express, expose and understand the semantics of the various systems. This variety has lead to, so-called, the *semantic heterogeneity* [16, 88, 46, 53].

Ontologies have gained a great attention in research as well as in industry for enabling knowledge representation and sharing. An ontology is a structure representation of critical knowledge that enables different systems sharing this knowledge to communicate meaningfully. Ontologies are considered an appropriate answer to the problem of semantic heterogeneity [16]. Although the use of ontologies may facilitate semantic interoperability, multiple users or organizations are likely to declare their own knowledge ontology (domain ontology) for describing and annotating their shared documents. Accordingly, many domain-ontology describing the same domain coexist independently created by different users. The proliferation of various domain ontologies has introduced more semantic heterogeneity [37, 16].

In response to the generalized heterogeneity on the growing amount of published ontologies on the Web. In the last two decades a specific research field has emerged, the so-called *Ontology Matching*. Ontology matching studies the ways to automatically establish a semantic relationships (correspondences) between two (or more) ontologies entities [37]. Ontology matching enables ontologies to interoperate. However, discovering (automatically, or even manually) such correspondences between different ontologies is a complex task, deep reasons of heterogeneity between ontologies to be matched are not explicitly known by machines (and for human to some extent) as explained before.

In general, matching methods are combinations of individual (atomic) matching techniques [102] that can be divided in four categories based on which kind of data the matching techniques work on [103]: terminological techniques, structural techniques, instance-based(extensional) techniques, and logical reasoning(semantic) techniques.

The *terminological* techniques (or, in general element-level techniques [37]) refers to the *string-based* and *linguistic-based* techniques that find correspondences between the ontologies textual entities descriptions and labels. String-based metrics take advantage of similar characters from two strings, whereas, linguistic-based metrics compare the meaning of strings. The underlying idea is that the more similar are two entities strings, the more they are likely referring to the same concept. Various string-based techniques were proposed to compute this similarity; one can simply compute the longest size of common substrings (prefixes, suffixes), or more sophisticated ones such as edit distance (e.g. Levenstein distance, Monger-Elkan distance, Jaro-Winkler distance), different string matching algorithms can be used here, for more details on string similarity methods you can refer to [24]. Another techniques leverage on linguistic tools as a pre-processing phase before the string-based comparison, making use of various NLP techniques (e.g., tokenization, lemmatization and stemming) in order to exploit their morphological properties. Language-based resources (like common knowledge, domain specific thesauri, lexical ontologies, or dictionaries) also introduced to bridge the gap between a syntactic information and its meanings. For instance, WordNet [40] gives all the senses of a given word (called synsets), and provides a directed relation graph between the synsets that represents the semantic relations between synsets. A comprehensive details on using WordNet for ontology matching can be consulted in [71].

The *structural-level* techniques [37] make use of the ontologies structural presentation. The structural-based approaches consider the ontology as a graph whose nodes represent the ontological entities and the edges are labeled with relation names. The problem of matching ontologies is viewed as a problem of matching graphs. The underlying assumption of is based on the fact that the similarity between two entities on two respective graphs impact the similarities between the respective neighbour entities in each graph, this idea can be grounded in many several ways; by comparing the nodes (entities) children, leaves, or comparing entities in the transitive closure, among others.

The basic idea of *instance-based* (*extensional-based*) mapping techniques is based on the analysis of statistics or distributions of class extensions, the more common instances of two concepts is, the more there are likely to denote the same concept [59]. *Instance-based* techniques can also rely on the instances properties or descriptions. Instances analysis can be exploited to compute similarities score between classes or to train classifiers for machine learning methods.

The *logical* reasoning (model, or semantic) based techniques exploit the semantic interpretation of the input ontologies and apply methods like propositional satisfiability (SAT) or description logics reasoning techniques to check the logical consistency of the candidate correspondences returned by a previous steps, or to deduce other correspondences from the previously discovered ones.

In [81] Mochol proposed a deeper classification of the matching systems taking in account several dimensions. Those dimensions are: input, approach, output, usage, documentation, and cost characteristics. More on classifications of matching methods can be found in [17, 33, 37].

As outlined before, *ontology matching* is a solution to the semantic heterogeneity problem. The last two decades have witnessed a wide range of ontology matching methods which have been successfully developed and evaluated in OAEI contests since 2004 [104]. Several recent surveys [21, 103] and books [37, 10] have been written on this field, also several conferences and workshops (e.g., Ontology Matching[1]) have specifically tackled this topic as well. The area is very active and has attracted a lot of attention last years, but it is far from being resolved [103, 90].

The majority of the proposed matching techniques in these systems mainly focused on mapping between ontologies that are lexicalized in the same natural language (so-called, *Mono-lingual Ontology Matching*, MOM). Moreover, methods developed for MOM systems cannot directly access the semantic information when ontologies are expressed in different natural languages. Though, there is a need for a methods that automatically reconcile information when ontologies are *lexicalized* in different natural languages [107, 46].

Recently, a notable efforts [34, 108, 44] were introduced in order to overcome, the language barriers; the problem of matching two ontologies that use more than one language each, at the same time they share (at least one) the same languages (so-called, *Multi-Lingual Ontology Matching*, MLOM). A specific case, is when the two

[1]ontologymatching.org

ontologies do not share any languages to be matched (so-called, *Cross-Language Ontology Matching*, CLOM) [108].

In spite of theses efforts, which I discusses more in details in the following sections, I also found that the semantic nature of cross-language mappings that cross-language ontology matching methods are expected to find has not been sufficiently investigated [6, 43]. Next, I describe the semantics of mappings in the monolingual ontology matching. Then, research efforts to date that aim to tackle the cross-lingual ontology mapping issue are discussed.

3.3 Mapping Semantic

Ontology mapping which can be seen as an expression that establishes relations between elements of two (or more) heterogeneous ontologies, a crisp mapping tell us that a certain concept are related to other concept in different ontology, and the type of relations are typically set theoretical relation $\{\equiv, \sqsubseteq, or \perp\}$, while the weighted mapping in addition associate a number (weight) to those relations. An interesting approach presented in [9] provides a formal semantic of weighted mapping between different ontologies, based on a classification interpretation of mappings, that is, a two concepts are said to by *extensionally equivalent* if the set of objects classified under one concepts can be also (re-)classified under the second concept.

Atencia et al. [9] approach provides a formal semantic of weighted mapping between *logically founded* ontologies, that give the notion of *logical consequences* of weighted mappings that allows to define a set of inference rules to derive a mapping from a set of existing mappings.

> *"..based on a classification interpretation of mappings: if O1 and O2 are two ontologies used to classify a common set X, then mappings between O1 and O2 are interpreted to encode how elements of X classified in the concepts of O1 are re-classified in the concepts of O2, and weights are interpreted to measure how precise and complete re-classifications are"* [9].

Atencia et al. represent a formal semantics for interpreting a confidence value (weight mapping) associated with a mapping. Atencia et al. approach relies on a classification interpretation of mappings which takes inspiration from the family of extensional based approaches (for more details on this see [37]) used in ontology matching techniques. Atencia et al take advantage of precision, recall, and F-measures, as they are used in the context of classification tasks, in their formalization of the weight mapping relation (subsumptions ($\sqsubseteq, \sqsupseteq$) and equivalence ($\equiv$)) that associate mappings to a closed subinterval $[a, b]$, where a and b are real numbers in the unit interval $[0, 1]$ respectively define the lower and upper bound that precision and recall fall in.

Intuitively speaking, suppose I have two ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$. Ontology $\mathcal{O}_1$ is used to classify the set of elements $\{x_1, ..., x_{10}\}$, and suppose the same elements are reclassified in ontology $\mathcal{O}_2$. I can measure the values of the set theoretical mapping by counting the classified elements. For example, suppose that the elements $\{x_1, ..., x_{10}\}$, classified under the concepts $C \in \mathcal{O}_1$ and $D \in \mathcal{O}_2$, then I can say that concept C and D are equivalent with a value 1.0 ($< C, D, \equiv, 1.0 >$). Similarly if the elements $\{x_1, ..., x_5\}$ classified under the concept $H \in \mathcal{O}_1$ then I can have a subsumption relation between H and D with a value 0.5 ($< H, D, \sqsubseteq, 0.5 >$). Atencia et al. define the *weighted mapping* as an expression that represents the set theoretical relation between two concepts belong to two different ontologies by associating those relations with a closed subinterval of [0,1].

Definition 3.1: **Weighted Mapping**,
Given a two ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$, a weighted mapping from $\mathcal{O}_1$ to $\mathcal{O}_2$ is a quadruple: *Weighed Mapping* $:= \langle C; D; r; [a, b]\rangle$, where $C \in \mathcal{O}_1$ and $D \in \mathcal{O}_2$, $r \in \{\sqsubseteq, \equiv, \sqsupseteq, \perp\}$, and a, b are real numbers in the unit interval $[0, 1]$.

3.3.1 Classificational Semantics for Weighted Mappings

Logical Semantics
Following the standard model-theoretic logical semantics based on interpreting classes as sets: an interpretation $\mathcal{I}$ is a pair $\mathcal{I} =< \triangle^{\mathcal{I}}, \cdot^{\mathcal{I}} >$ where $\triangle^{\mathcal{I}}$ is a non-empty set, called domain of interpretation $\mathcal{I}$, and $\cdot^{\mathcal{I}}$is a function that interprets each concept (class) $c \in C$ as a non empty subset of $\triangle^{\mathcal{I}}$, and each instance identifier ($x \in X$) as an element of $\triangle^{\mathcal{I}}$.

Given an ontology $\mathcal{O}$, let C be a set of concepts, R a set of relations, and X a set of shared objects. *Then,* $c^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}}$ for $c \in C$, $r^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}} \times \triangle^{\mathcal{I}}$ for $r \in R$, and $x \in \triangle^{\mathcal{I}}$ for $x \in X$.

Assume that the concepts of two ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$, are used to classify a common set of elements X. Then the mappings between concepts in $\mathcal{O}_1$ and $\mathcal{O}_2$ encode how the elements of X classified under the concepts of $\mathcal{O}_1$ are re-classified in the concepts of $\mathcal{O}_2$, and the weights encode how precise and complete these re-classifications are.

Atencia et al.[9]:*Let $X = \{x_1, ..., x_n\}$ be a non-empty finite set of fresh constants not occurring in L($\mathcal{O}_1$) or L($\mathcal{O}_2$).The set X is meant to represent the set of shared items classified by concepts of the ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$. A classification of X in $\mathcal{O}_1$ is specified by virtue of an interpretation $\mathcal{I}_1$ of $\mathcal{O}_1$ extended with the elements of X as follows.*

Weighted Mapping Semantic

Let C be a concept of $\mathcal{O}_1$ and x_k a fresh constant of X; I define X as a *shared context* (domain) of the mapping. Then, x_k is classified under C according to $\mathcal{I}_1$ if $x_k^{\mathcal{I}_1} \in C^{\mathcal{I}_1}$. Then, the set $C_X^{\mathcal{I}_1} = \{x \in X | x^{\mathcal{I}_1} \in C^{\mathcal{I}_1}\}$ represents the subset of items of X classified under C according to $\mathcal{I}_1$.

Note that $C_X^{\mathcal{I}_1}$ is a subset of X ($C_X^{\mathcal{I}_1} \subseteq X$) whereas $C^{\mathcal{I}_1}$ is a subset of the domain of the interpretation $\mathcal{I}_1$ ($C^{\mathcal{I}_1} \subseteq (\triangle^{\mathcal{I}_1}$). In addition, $C_X^{\mathcal{I}_1}$ is always a finite set while $C^{\mathcal{I}_1}$ may be infinite.
Let $\mathcal{I}_1$and $\mathcal{I}_2$ be interpretations of $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively, and let C and D be the concepts of $\mathcal{O}_1$ and $\mathcal{O}_2$, occurring in the correspondence $\langle C, D, r, [0, 1]\rangle$. The sets $C_X^{\mathcal{I}_1}$ and $D_X^{\mathcal{I}_2}$can be compared as they are both subsets of X which represents the sets of items of X classified under C according to $\mathcal{I}_1$ and under D according to $\mathcal{I}_2$, respectively. Then the different types of mappings $\langle C, D, r, [0, 1]\rangle$ obtained by looking at the different $r \in \{\sqsubseteq, \sqsupseteq, \equiv, \perp\}$.

Intuitively, following the classification tasks, the mapping $\langle C, D, \sqsubseteq, [0, 1]\rangle$ is used to express that any item in X which is classified under C according to $\mathcal{I}_1$ is (re-)classified under D according to $\mathcal{I}_2$. Where the confidence level interval [0,1] (the weighted mapping, [9]) can be seen as the recall of $C_X^{\mathcal{I}_1}$ w.r.t $D_X^{\mathcal{I}_2}$.

$$\mathrm{R}(C_X^{\mathcal{I}_1}, D_X^{\mathcal{I}_2}) = \frac{|C_X^{\mathcal{I}_1} \bigcap D_X^{\mathcal{I}_2}|}{|C_X^{\mathcal{I}_1}|}$$

In the same way, the mapping $\langle C, D, \sqsupseteq, [0,1]\rangle$ is used to express the fact that the fraction of items of X classified by D according to $\mathcal{I}_2$ which are (re-) classified under C according to $\mathcal{I}_1$. Where the confidence level (weighted mapping) can be seen as the precision of $D_X^{\mathcal{I}_2}$ w.r.t $C_X^{\mathcal{I}_1}$.

$$\mathrm{P}(C_X^{\mathcal{I}_1}, D_X^{\mathcal{I}_2}) = \frac{|C_X^{\mathcal{I}_1} \bigcap D_X^{\mathcal{I}_2}|}{|D_X^{\mathcal{I}_2}|}$$

By keeping the parallelism with classification systems, the natural way to interpret the confidence level (weighted mapping) of the equivalent relation that aligns two concepts C and D, $\langle C, D, \equiv, [0,1]\rangle$, is by means of the F-measure, which is the harmonic mean of precision and recall. Typically the F-measure used to evaluate the global quality of a classifier, The F-measure of $C_X^{\mathcal{I}_1}$ and $D_X^{\mathcal{I}_2}$ is defined as

$$\mathrm{F}(C_X^{\mathcal{I}_1}, D_X^{\mathcal{I}_2}) = 2\frac{|C_X^{\mathcal{I}_1} \bigcap D_X^{\mathcal{I}_2}|}{|C_X^{\mathcal{I}_1}|+|D_X^{\mathcal{I}_2}|}$$

An interesting point in Atencia et al. weighted mapping definition is the use of ranges of scores [a, b] for subsumption relation that interpreted as the precision $\langle C, D, \sqsubseteq, [a,b]\rangle$, and recall $\langle C, D, \sqsupseteq, [a,b]\rangle$, by this I can define the equivalence relation as a conjunction of the two subsumption relations. This in particular gives the notion of logical consequences of weighted mappings that allows to define a set of inference rules to derive a mapping from a set of existing mapping.

For instance, if I have a weighted mappings $\langle C, D, \sqsubseteq, [c,d]\rangle$ and $\langle C, D, \sqsupseteq, [e,f]\rangle$, then I can derive the equivalence weighted mapping $\langle C, D, \equiv, [v,w]\rangle$ with $v = min(c,d)$ and $w = max(d,f)$. Notice that, if I consider the usual definition of equivalence in DLs in terms of subsumption: $\langle C \equiv D\rangle$ iff $\langle C \sqsubseteq D\rangle$ and $\langle C \sqsupseteq D\rangle$, when dealing with single values for precision($\sqsubseteq$) and recall($\sqsupseteq$) instead of intervals, it is usually impossible to combine them into a single value by simple conjunction [9].
Nevertheless, generally ontology matchers used to return a single confidence level value, for instance, n. Accordingly, to represent the value n by means of the weighted mapping interval [a, b], the authors in [9] suggest to use a pointwise interval, I can assume that $a = b$, then n=[a, a]. Thus, I can simply present the mapping relation as $\langle C, D, r, n\rangle$.

Figure 3.1, demonstrates the extensional meaning between two concepts C and D of the ontology $\mathcal{O}_1$ and ontology $\mathcal{O}_2$ respectively, based on the classification

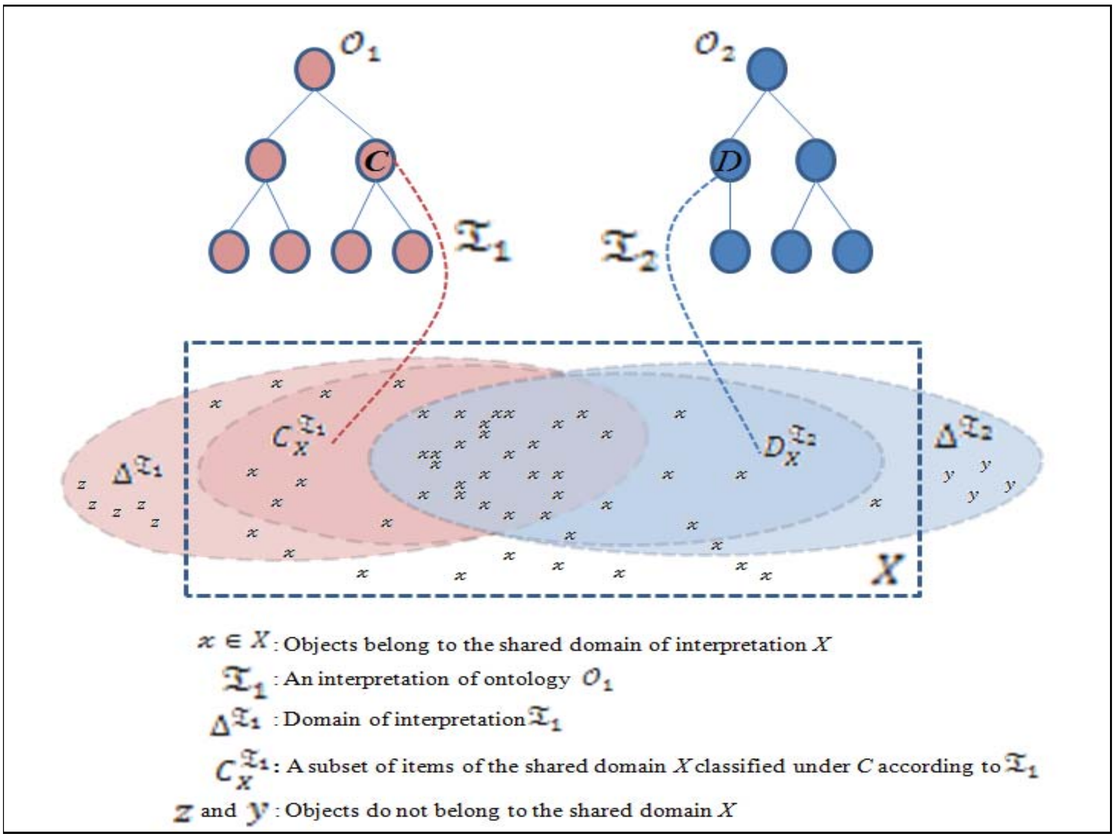


FIGURE 3.1: The extensional meaning of concepts

based mapping approach. $\mathcal{I}_1$ and $\mathcal{I}_2$ represent an interpretation of $\mathcal{O}_1$, and $\mathcal{O}_2$, respectively. And $\triangle^{\mathcal{I}_1}$ and $\triangle^{\mathcal{I}_2}$ represent the domain of interpretation of $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively.

The set of element X represent the shard domain of interpretation between the two ontologies. Note that, the elements z and y that belong to $\triangle^{\mathcal{I}_1}$ and $\triangle^{\mathcal{I}_2}$, respectively, do not belongs to the shared domain X. The sets $C_X^{\mathcal{I}_1}$ and $D_X^{\mathcal{I}_2}$ represents the subsets of items of the shared domain X classified under C according to $\mathcal{I}_1$ and under D according to $\mathcal{I}_2$, respectively.

Challenges and Open Issues

In the original definition of the extensional meaning of a mapping using the classification based approach, that Atencia et al. proposed, assumes a logical interpretation as a concept denoted as class of instances in an interpretation domain. The extensional meaning of a concept is interpreted as a subset of objects in a shared domain of interpretation provided by $(\mathcal{I}_1 \bigcup \mathcal{I}_2 \bigcup X)$.

In the logical domain the interpretation of classification is a concept classifies an individual, where these individuals are a members of a class. The extensional meaning in this case cannot be directly adapted to ontologies that do not have such logical interpretation of classification. For instance, when I annotate a document I can consider the concept as classifying an object, but the interpretation of classification is different, in this case saying that a concept classifies an object, means that the concept is the topic of the document. While if I consider a text where I have several terms and I want to provide a disambiguation the meaning of the term I can classify a term with a concept, saying that the sense of the term is the associated concept.

One can extend the extensional interpretation of mapping in the logical domain for other type of ontologies using different way of interpretation of extension and different interpretation of the notion of classification of an instances with a concept. Besides that, Atencia et al. classification approach considers a finite set of objects belonging to the shared context of interpretation, while if we consider a generic ontologies representing concepts lexicalized in languages that spoken by very large community, the shared context (or domain) of interpretation of the mapping problem might be very large or even infinite, because some concepts refer to objects that might have infinite extension. The question here is, what is the impact on this formalism if we consider an infinite set of objects in the shared context.

Moreover, the proposed approach represents a semantic mapping between two ontologies belong to the same type of interpretation, logical ontologies in this case. An interested research direction might be the study of mapping two ontologies which are interpreted in different way (cross-ontology interpretation), i.e., can such semantic be extend to map mixed interpretations, e.g., lexical and logical ontologies.

I believe that such approach can fits the CLOM problem. However, in ordered to adopt to a border notion of ontology, which encompasses lexical ontologies and logical ontologies, the classification based mapping approach presented in [9]. I need to extend this definition using the classification based approach *independently* of the interpretation of classification and the type objects that can be classified under the concept, as well as to consider the concept *lexicalization* in the classification based approach which is a fundamental aspect used by ontology matchers and central point toward extending such approach for the cross-lingual matching problem.

In Chapter 4, I discus a classification-based semantic for cross-lingual mapping settings.

3.4 Cross-Lingual Ontology Mapping

The majority of *ontology mapping* methods proposed in the literature have addressed the problem of mapping ontological resources lexicalized in the same natural language, called *mono-lingual ontology mapping.* Since mono-lingual matching systems cannot directly access semantic information when ontologies are lexicalized in different natural languages [44], techniques to reconcile ontologies lexicalized in different natural languages have been proposed [46, 111].

Translation-driven approaches have been used to overcome the natural language barriers by transforming a cross-lingual mapping problem into a mono-lingual one [44]. Different multilingual lexical resources have been used to perform the translation tasks, including manual translations, machine translation tools, and bilingual dictionaries built from Web-based multilingual resources. For a rich classification and comparison of cross-lingual mapping systems I refer to [111].

Liang and Sini [68] *manually* mapped the English thesaurus AGROVOC[2] to a Chinese thesaurus CAT[3]. The mappings generated by such approaches are likely to be accurate and reliable. However, this can be a time and resource consuming process specially for maintaining large and complex ontologies.

Machine translation tools are widely adopted for cross-lingual ontology mapping. Sphor et al., [107] translate the ontology labels into a pivot language (English) using the machine translation tool Bing[4]. Then, they define a feature vector based on a combination of string-based and structural-based similarity metrics and learn a matching function using a support vector machine. Like other approaches based on supervised machine learning algorithms, their approach has the disadvantage of requiring a significant number of training samples and well-designed features to achieve good performance. Fu et al., [44] translate ontology labels using Google Translate, and then match these translated labels by combining different similarity

[2] http://aims.fao.org/website/AGROVOC-Thesaurus/sub
[3] http://www.ciard.net/partners/labof-chinese-agricultural-ontology-services
[4] http://www.bing.com/translator

measures. Their approach leverages structural information about the ontology concepts by considering their neighbours in the matching process. Other approaches have been proposed that also apply string-based, lexical and structural matching methods to ontology labels translated with machine translation tools, like Google Translate or Bing [38, 64, 32].

Multilingual knowledge resources available on the web have been also exploited to translate concepts' labels [57]. *Wiktionary*[5] was used to generate translations to match English and French ontologies [70]. First, a bilingual English-French lexicon is built using *Wiktionary* and is used to translate the labels of the ontologies. Then, the monolingual ontology matching system *COMS* is used [69]. *COMS* uses a set of string-based, lexical and structural matching techniques to find the appropriate mappings. A similar approach uses *Wikipedia* inter-lingual links to retrieve candidate matches for source concepts [15, 51]. However, when used alone, Wiktionary and Wikipedia inter-lingual links may have limited coverage, in particular for resource-poor languages.

In spite of these efforts, cross-lingual mapping systems still perform significantly worse than mono-lingual mapping systems according to recent results in the OAEI contest [104], which suggest that cross-lingual ontology mapping is still a very challenging problem [111]. The datasets used to evaluate cross-lingual mapping in the OAEI, i.e., the datasets in the multifarm track [77], consist of alignments established between axiomatic ontologies of relatively small size and specific to the domain of conference organization. Since in this thesis I we want to investigate translations obtained with different multilingual lexical resources at a large scale and not in a specific domain, I decided to use different and larger gold standards in the experiments.

3.5 Enrichment of Multilingual Knowledge Resources

The enrichment of multilingual knowledge resources is related to cross-lingual ontology mapping for several reasons. First, multilingual knowledge resources can be used as sources of translations in cross-lingual ontology mapping approaches.

[5] https://www.wiktionary.org

Second, the wordnets mapped to the English WordNet that I use as gold standards throughout this thesis (Section 2.5.1) are multilingual knowledge resources, because their mappings represent inter-lingual links between concepts. Third, the two most frequently adopted approaches to enrich multilingual knowledge resources are based either on mapping concepts lexicalized in different languages or on translating the concepts lexicalizations. Since I want to analyze the correctness and coverage of translations of ontology concepts, the analyses findings are relevant also to approaches that intend to use these translations and ontology mapping methods to enrich multilingual knowledge resources.

Several multilingual wordnets (lexical ontologies) were developed by manually or automatically translating concepts of the English WordNet into new languages [93, 115, 112, 49, 110]. The *expand* and *merge* models [115] are the main approaches used in the development of multilingual wordnets. In the merge model, synsets of a pre-existing resource in one language (e.g., a thesaurus, or even an unstructured lexical resource like a dictionary) are aligned to the most equivalent synset in English. In the expand model, English synsets are translated into the respective languages. The main advantage of these two approaches is to avoid the expensive manual elaboration of the *semantic hierarchy* in new languages. The English WordNet [40] hierarchy is used as reference for all wordnets. Moreover, any ontology that is built following these approaches is also automatically mapped to the English WordNet.

In several wordnets, the English concepts were *manually* translated by human lexicographers using external lexical resources such as dictionaries, thesauri and taxonomies. This approach has been applied to build for example the Arabic wordnet [97], the Italian wordnet [93], the Spanish wordnet [49] and the core of the Slovene wordnet, all used in the experiments. However, the manual approach to construct ontologies that aim to cover natural languages' lexicons is often an effort-intensive and time-consuming task [30]. *Automatic* approaches have been therefore proposed to reduce the lexicographers' workload.

Parallel corpora have been used in building wordnets for languages other than English. The basic assumption underlying these methods is that the translations of words in real texts offer insights into their semantics [96]. The Slovene wordnet was enriched using word alignments generated by a sentence-aligned multilingual corpus [41]. The wordnet has been further extended using bilingual dictionaries and inter-lingual links in Wikipedia. A similar approach is also followed in building the

French wordnet [99]. The monosemous words in the English WordNet were automatically translated using bilingual French-English dictionaries built from various multilingual resources, such as Wikipedia inter-lingual links, Wiktionary, Wikispecies[6], and the EUROVOC thesaurus[7].

Sentence-aligned parallel corpora may not be available for all pair of all natural languages. In addition, specific tools are needed to perform sentence and/or word alignment across the corpora, and the bilingual dictionaries extracted from these corpora are biased towards the domains they cover. To overcome these limitations, in the Macedonian wordnet [101], *a machine translation tool* has been used to create parallel corpora. Monosemous English words were directly translated using a bilingual English-Macedonian dictionary. For polysemous words, an English sense-tagged corpus (the English WordNet tagged glosses[8]) was automatically translated into Macedonian using Google Translate.

A supervised method to automatically enrich English synsets with lexicalizations in other languages was also proposed [31]. This method learns to determine the best translation for English synsets by taking into account bilingual dictionaries, structural information in the English WordNet, and corpus frequency information.

Other approaches to enrich multilingual knowledge resources have been proposed to build the Universal WordNet (UWN) [30], WikiNet [84], and BabelNet [86], which integrate multilingual encyclopedic knowledge from Wikipedia with the English WordNet. In this paper, I focus on BabelNet, the largest multilingual knowledge resource as of today, and use it in the study to build bilingual dictionaries that I use for translation (explained in Section 2.4.1). A comprehensive comparison amongst the afore-mentioned three resources can be found in [86].

BabelNet [86] has been built by integrating the English WordNet with Wikipedia. These two resources have been mapped using an unsupervised approach. As a result, BabelNet covers approximately the 83% of WordNet's nominal synsets. Synsets from the English WordNet cover in particular (but not only) classes of objects, e.g., "University"[9], while Wikipedia entries cover in particular (but not only) named entities, e.g., "University of Milano-Bicocca"[10,11]. Synsets from the English

[6] https://species.wikimedia.org
[7] http://europa.eu/eurovoc
[8] http://wordnet.princeton.edu/glosstag.shtml
[9] http://wordnetweb.princeton.edu/perl/webwn?s=university
[10] https://en.wikipedia.org/wiki/University_of_Milano-Bicocca
[11] http://babelnet.org/synset?word=University_of_Milano-Bicocca

WordNet and other BabelNet entries are *enriched* with lexicalizations in other languages using a variety of lexical resources. A first set of lexicalizations in languages other than English are obtained by using inter-lingual links of Wikipedia. Synsets for which Wikipedia entries cannot be found have been enriched using automatic translations of English senses-tagged sentences, extracted from Wikipedia and the *SemCor* corpus [80]. The most *frequent* translation in a given language is detected and included as a variant lexicalization in this language; this approach was named *context-translation*. Translations of *monosemous* English words have been collected using Google Translate and directly included in the expanded lexicalizations; this approach was named *contextless-translation*. Observe that contextless translations are based on an heuristics, i.e., that monosemous words are correctly translated (also referred to as monosemous word heuristics). The *core* of BabelNet consists of the lexicalizations obtained with these approaches, also named *BabelNet synsets*. Later, BabelNet synsets' lexicalizations are expanded with more multilingual lexical resources[12]: Wiktionary, WikiData[13], OmegaWiki[14], and several wordnets that are mapped to the English wordnet, which are available through the Open Multilingual Wordnet[15][14].

BabelNet lexicalizations (synsets) have been evaluated against manually mapped wordnets, which I also use in the experiments in this thesis as gold standards. They also performed a manual evaluation with a randomly sampled set of concepts. A limit of their evaluation consists in not making explicit if the sampled senses uniformly cover polysemous and monosemous senses. Otherwise this distinction is important to evaluate different translations, also because a vast number of translations have been obtained using the contextless approach, which is based on the monosemous word heuristics. In the experiments (Section 5.4) I specifically analyze the effectiveness of the monosemous word heuristics in the context of ontology mapping.

I observed that the expand model was used more substantially than the merge model in approaches to automate the enrichment of multilingual wordnets and knowledge resources. One may attempt to enrich an existing wordnet via the merge approach by mapping an unstructured or a weakly structured lexicon, e.g., a dictionary, to a structured reference ontology, e.g., the English WordNet. For

[12]http://babelnet.org/stats
[13]http://www.wikidata.org
[14]http://www.omegawiki.org
[15]http://compling.hss.ntu.edu.sg/omw/

example, in the Arabic Ontology Project [60, 6], the authors plan to use this approach to extend a core ontology manually created and mapped to the English WordNet. However, the mapping task incorporated in this approach is particularly challenging [1]: the lack of semantic relations between the synsets of an unstructured lexicon makes it difficult to disambiguate their meaning during the translation and the matching steps [103, 111]. An effective cross-lingual ontology mapping method can support the application of the merge model at large scale, thus supporting the construction and enrichment of multilingual knowledge resources. For example, In Chapter 6 I propose an approach, despite the difficulty of the task, returns multilingual concept lexicalizations richer than the ones that can be obtained by automatically translating the concepts labels [3].

3.6 Cross-Lingual Word Sense Disambiguation

Cross-lingual ontology mapping is also related to the Cross-lingual Word Sense Disambiguation problem (*CL-WSD*), which has been studied in the recent past and addressed in SemEval 2010 and 2013 challenges [36, 67]. The goal of *CL-WSD* is to predict semantically correct translations for ambiguous words in context [96].

In *CL-WSD*, the lexical disambiguation task is performed as a word translation task, called *lexical substitution task* [75]. Given a source word in a sentence (e.g., an Italian word), the system tries to translate the word into a different language (e.g., English). The translation is considered to be correct if it preserves the sense that the word has in its context also in the target language.

Most of *CL-WSD* systems rely on *parallel corpora* [45, 96, 8], including those which exploit existing multilingual wordnets [58]. However, the success and coverage of these methods highly depends on the nature of the parallel corpora and on the way the extracted information is used to select the appropriate senses. Corpora are known to have domain-orientated coverage, i.e., fine-grained senses for different domains might not be found in specific parallel corpora [85]. More importantly, parallel corpora may not be available for language couples or for specific domains [8, 101].

One fundamental difference between the CL-WSD task and cross-lingual ontology mapping is that in CL-WSD a context is always available and defined by the sentence a word occurs in. In cross-lingual ontology mapping the context can be defined by the neighbours of a translated concept, may be limited [82, 117], or may not be even available, e.g., when an unstructured lexicon is matched against a structured ontology [60, 6].

3.7 Summary

In spite of the fact that, most of the approaches to cross-lingual ontology mapping is based on transforming a cross-lingual matching problem into a monolingual one by leveraging automatic machine translation tools [111], few efforts have been dedicated to systematically study the effectiveness of automatic machine translations in cross-lingual ontology mapping.

Fu et al. [43] studied some limitations of this approach, in particular, to what extent the inadequate translations can introduce "noise" into the subsequent mapping step, where matches may be neglected by matching techniques that depend only on the discovery of lexical-based similarities. They performed two case studies that examined mappings of independent, domain specific, and small-scale ontologies that are labeled in English and Chinese: the Semantic Web for Research Communities ontology and the ISWC ontology. The ontologies have not been lexically enriched. Fu et al. [43] classified the translation errors introduced by the machine translation tools into three main categories. *Inadequate* translation, when the returned translation is more specific/generic concept instead of the intended concept presented in the original ontology; *Synonymic* translation, when the translation result of a label is correct, however it is different with the one that was used by the target ontology; and *Incorrect* translation, when the translation is simply wrong. In addition, the study showed that translating ontology labels in isolation leads to poorly translated ontologies which then yields low-quality matching results, thus, label translations should be conducted within context. The context is characterized by the surrounding ontology concepts.

Sphor et al. [107] leveraged the multilingual evidences when ontologies are labeled with more than one languages (called Multilingual Ontology Mapping). They concluded that the best approach is to translate the labels in the source ontology to all languages available in the target ontology when trying to match a monolingual source ontology to a multilingual target ontology. However, this can be applicable only when multilingual labels are available, which is not the case for large number of cross-lingual mapping scenarios. In addition, to obtain better translations Sphor et al. [107] study suggested that translating the source and target ontologies' labels into a pivot language, to some extent, can improve the quality of the translation. However, the authors stated that, further evidence from several language pairs are needed in order to be support this claim, as the translation quality of machine translation systems greatly depends on the pair of languages considered.

In this thesis, In Chapter 5, I analyze the effectiveness of automatic translations for cross-lingual concept mapping using large scale, general domain, and lexically rich ontologies (wordnets). The ontologies used in though out the studies cover four different families of languages besides English. I study effectiveness of translation resources by conducting a large number of experiments and in relation to both the candidate match retrieval and the mapping selection tasks. Overall, I believe that none of previous work on cross-lingual ontology mapping provided such a systematic study, if compared in terms of scale (size of the considered concepts), number of considered languages, and detail of the analysis (concept categorization).

The analyses discussed in this thesis can be also related to the studies on automatic translation strategies conducted to evaluate BebelNet, which is one of the two translation resources considered in the conducted study. In this work, I quantitatively evaluate the correctness and the coverage of the translation strategies used in BabelNet as means to support cross-lingual mapping tasks (using manually established mappings between wordnets for comparison; see Section 5.4.1). The studies conducted to evaluate BabelNet were aimed, instead, at evaluating their translation strategies as means to collect multilingual lexicalizations for the concepts. Several of the measures used in the study are borrowed from BabelNet (but adapted to the proposed formalization), but I also introduce a new measure and to assess synonym word coverage (see section 5.3 and Section 5.4.2). In addition, coverage and correctness of automatic translations in the study are evaluated by considering different categories of synset (defined in Section 5.2). Finally, I analyze the effectiveness of the monosemous word heuristic, which is used in several matching systems and in BabelNet (contextless translation).

Chapter 4

Cross-Lingual Mapping Semantics

4.1 Chapter Overview

This chapter presents an overview of the semantics of cross-lingual mapping based on an extensional model. In Section 4.2, I present the foundation of the proposed semantics for cross-language mapping. In other words, I define the semantics of a correspondence for cross-language ontology matching, taking into account the lexicalization in the definition of mapping. Further, in Section 4.3, I provide an overview of an experimental setting, through which the proposed mapping semantics can be evaluated and a gold standard dataset can be generated. In Section 4.3.2, I discuss the limitations concerning the proposed validation approach, and I propose different settings to overcome the challenges. Finally, in Section 4.4, I conclude and outline the next steps.

The scientific accomplishment presented in this chapter have been published in [6, 2, 62].

4.2 Classification-based Interpretation in Cross-Lingual Mappings for Lexical Ontologies

In the context of cross-lingual mapping, languages barrier has been approached by transforming a cross-lingual mapping problem into a mono-lingual mapping one by leveraging translation tools [111]. However, the cultural-linguistic barriers still need more efforts in terms of the mapping process and techniques [22, 46], as well as to formally define the mappings semantic that align concepts lexicalized in different natural languages. Without loss of generality, if two concepts are lexicalized in different languages, then they are considered equivalent if they express the same meaning in a given context (i.e., same concept). If two language communities (the majority of language speakers) share the same understanding for a given concept, irrespective of the lexical notation being used (language), then they refer to the instances that are classified under this concept[1]. The goal is to provide a formal definition of the cross-lingual mapping problem, mainly *to define what a correspondence is, and how to represent correspondences in cross-lingual mapping problem*, that is, to define the semantics of the correspondence considering lexicalization in the definition of the mapping.

To achieve this goal, I need to define a formal interpretation (i.e., formal semantics) of the correspondence in the cross-lingual mapping settings . For this, I extend the formal definitions of the classification-based semantic (discussed in Section 3.3), which I found suitable for this purpose.

The extension of concept is often used in many cross-lingual ontology matching strategies [37, 103, 90]; this extension is interpreted in different ways, e.g., instances classified under concepts, or even a document annotated with a concept. I believe this is a promising approach to provide a foundation to cross-lingual mapping semantics, and it makes sense to adopt such an approach that is based on the classification of different kind of objects with a concept to interpret the semantics of mapping. The classification-based approach defines semantic based on the use of the concepts and the classes for objects. I believe that this approach is useful in cross-lingual mapping case since it can also support the definition of translation between concepts in ontologies, which are lexicalized in different languages, based on the use of these concepts as classifiers.

At the same time, the adaptation of the classification-based semantic to a cross-lingual mapping setting is not trivial. The Atacia et al. [9] approach (explained in Section 3.3) does not explicitly introduce the lexicalization of the ontologies, which is important for it to be used by ontology matchers [37]. As a result, I provided a lexicalized version of the ontology matching problem mainly in the context of

[1]The context (or discourse) that a community of speakers shares in order to decide if these two terms (lexemes) refer to the same concept is "not only to explain what people say, but also how they say it. Lexical choice, syntax, and many other properties of the formal style of this speech are controlled by the parliamentary context" [113].

cross-lingual mapping based on classification interpretation of mappings, which I discuss in the upcoming section.

4.2.1 Formal Interpretation

I interpret the **classification task** as a task to establish whether an instance i is member of a class C, i.e., if i belongs to the extension of C. This extensional interpretation cannot be directly applied for ontologies that are not formally (logically) represented and interpreted in set theoretic semantics. If I consider a sentence and I want to disambiguate the meaning of the words in it, I can consider the **disambiguation task** as a form of classification, namely, the classification of a word as occurrence of a word sense in the sentence [85, 67].

I extend the notion of a mono-lingual mapping definition to the cross-lingua mapping one by considering the lexicalization of the ontology entities; I adapt the approach presented in [9] (explained in Section 3.3) that provides a formal interpretation of the semantics for the weighted ontology mapping based on the extension of concepts, in mono-lingual mapping settings.

Let $X = \{x_1, ..., x_n\}$ be a non empty finite set of instances constants, and let C_l be a concept lexicalized in language l; I say that instance x_n is classified under C_l according to the interpretation $\mathcal{I}_1$ if $x_n^{\mathcal{I}_1} \in C_l^{\mathcal{I}_1}$. Then, the set $C_l^{\mathcal{I}_1} = \{x \in X | x^{\mathcal{I}_1} \in C_l^{\mathcal{I}_1}\}$ represents the subset of instances belong to X classified under the lexicalized concept in a given language l, C_l, according to the interpretation $\mathcal{I}_1$. Note that $C_{X,l}^{\mathcal{I}_1} \subseteq X$ and $C_l^{\mathcal{I}_1} \subseteq \triangle^{\mathcal{I}_1}$, where $\triangle^{\mathcal{I}_1}$ is a non-empty set, called domain of interpretation $\mathcal{I}_1$.

Let $\mathcal{I}_1$ be interpretation of ontology $\mathcal{O}_1$ lexicalized in language l, and $\mathcal{I}_2$ be interpretation of ontology $\mathcal{O}_2$ lexicalized in language $l^{'}$. Let C and D be a lexicalized concepts of $\mathcal{O}_1$, and $\mathcal{O}_2$, respectively, occurring in the correspondence mapping $\langle C_l, D_{l'}, r, [a, b]\rangle$, where r is a relation that holds between C_l and $D_{l'}$ with a confidence level interval $[a, b]$. Then, the sets $C_{X,l}^{\mathcal{I}_1}$ and $D_{X,l'}^{\mathcal{I}_2}$ can be compared as they are both subsets of X which represent the sets of objects of X classified under the lexicalized concept C_l according to $\mathcal{I}_1$ and under the lexicalized concept $D_{l'}$ according to $\mathcal{I}_2$, respectively. Following the classification based mapping explained in Section 3.3, I interpret the confidence level of the extensional equivalent relation by means of F_1-measure, which is the harmonic mean of recall and precision.

$$\text{Recall: } \mathrm{R}(C^{\mathcal{I}_1}_{X,l}, D^{\mathcal{I}_2}_{l'}) = \frac{|C^{\mathcal{I}_1}_{X,l} \bigcap D^{\mathcal{I}_2}_{X,l'}|}{|C^{\mathcal{I}_1}_{X,l}|}$$

$$\text{Precision: } \mathrm{P}(C^{\mathcal{I}_1}_{X,l}, D^{\mathcal{I}_2}_{l'}) = \frac{|C^{\mathcal{I}_1}_{X,l} \bigcap D^{\mathcal{I}_2}_{X,l'}|}{|D^{\mathcal{I}_2}_{X,l'}|}$$

The F_1-measure of $C^{\mathcal{I}_1}_{X,l}$ and $D^{\mathcal{I}_2}_{X,l'}$ is defined as $\mathrm{F}(C^{\mathcal{I}_1}_{X,l}, D^{\mathcal{I}_2}_{X,l'}) = 2\frac{|C^{\mathcal{I}_1}_{X,l} \bigcap D^{\mathcal{I}_2}_{X,l'}|}{|C^{\mathcal{I}_1}_{X,l}|+|D^{\mathcal{I}_2}_{X,l'}|}$.

As introduced before, in Section 3.3, the weight mapping between two objects is represented by means of an interval $[a, b]$, while in general an ontology matching algorithm used to return a single confidence level value, for instance, n. Accordingly, to represent this value n by means of the weighted mapping interval $[a, b]$, a *pointwise interval* can be used, that is, I assume that $a = b$, then $n = [a, a]$. Thus, I can simply present the mapping relation as $\langle C_l, D_{l'}, r, n\rangle$.

Next, I provide an overview of an experimental setting, through which the proposed mapping semantics can be evaluated and a gold standard dataset can be generated. Such datasets can be used as reference alignments to assess the quality and to compare alternative cross-lingual mapping methods.

4.3 Experiment Design for Cross-lingual Mapping Validation

In this section, I present an experimental setting that is aimed at showing that the above described semantics can be used, in principle, to define cross-lingual ontology alignments by assigning classification tasks to bilingual speakers. As introduced before in Chapter 1, this is important because a gold standard is needed to comparatively evaluate alternative cross-language mapping methods and few high-quality gold standards are available at present, e.g., OAEI multifarm [77], which is a small and specific domain dataset.

Consider a corpus of sentences, where each sentence expresses a context and a word in the sentence represent the usage of a concept. If a majority of speakers (i.e., bilingual native speakers or lexicographers) can substitute two words, each belonging to a different language, in a sentence and both words indicate the same sense (meaning), then these words can be used interchangeably to refer to the same concept (word sense). I hypothesize that, if speakers can substitute two words in a given context, then these words are synonyms and give an equivalent meaning [40]. This is valid also for intra- and inter-lingual substitution, as concepts are independent of specific languages. I assume the above hypothesis, however, instead of considering the cross-language substitutability of words themselves, I consider the cross-lingual substitutability of meanings associated with these words, by referring to co-disambiguation (Definition 4.1) of words across ontologies in different languages, similar tasks have been also introduced in [67, 36, 75].

Definition 4.1: Co-disambiguation task (CL-WSD),
Let $WSD(w_i)$ be a function called Word Sense Disambiguation, such that w_i is an occurrence of the word w in a sentence S. WSD associates w_i with a sense in a lexicon (e.g., WordNet). Accordingly, I define a cross-lingual WSD function $CL{-}WSD_{[L_1,L_2]}(w_i)$, such that CL-WSD associates a word w_i in a language L_1 (where L_1 is the language used in sentence S) with a word sense (concept) in a lexicon lexicalized in another language L_2.

In another words, if the substitution of the words does not change the meaning of the context, then they are conceptually equivalent. In view of this, CL-WSD can be seen as a classifier, where the number of agreements among the bilingual speakers (lexicographers) expresses the confidence (i.e., the weight) of the mapping. Bilingual speakers perform the CL-WSD tasks, and the mapping between two word senses depends on a frequency-based function that measures the degree in which the two senses in two different languages co-disambiguate the same word sense in multiple contexts (sentences).

4.3.1 Experimental Settings

This section describes experimental settings in which the proposed semantic is used in order to have equivalent relation between concepts in different languages. Given a corpus of sentences in language L_1 (e.g., English), I find a word $w_{en,i}$ that appears in these sentences. I disambiguate each occurrence of $w_{en,i}$ with an English concept C_i; and I disambiguate each occurrence of $w_{en,i}$ with a concept D_i in language L_2 (e.g., Arabic). As a result of this operation I found two sets of distinct concepts $\bar{C}$ and $\bar{D}$ that have been used to disambiguate $w_{en,i}$ respectively in English and Arabic. For each $C_i \in \bar{C}$ I count the number of C_i that has been co-disambiguated with every $D_i \in \bar{D}$. The co-disambiguation fraction of the two

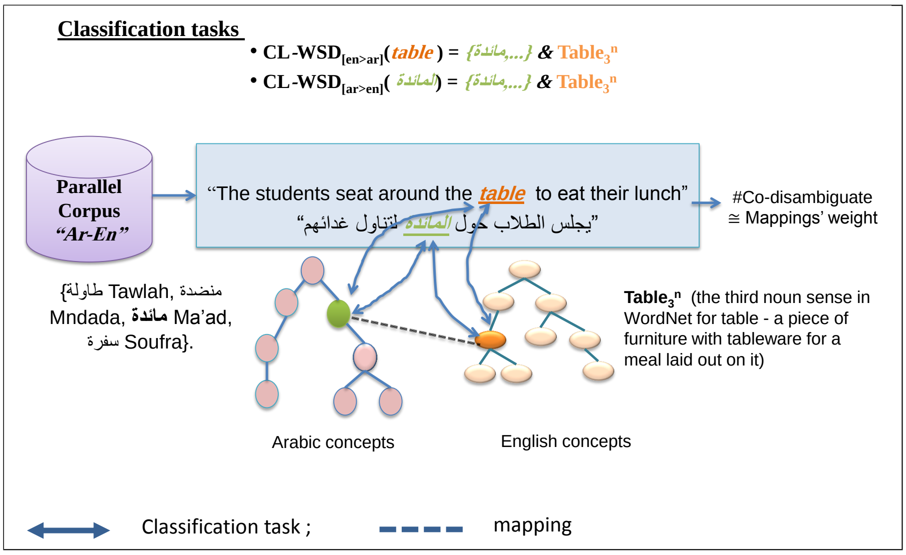


FIGURE 4.1: Example of Co-disambiguation task

concepts C and D represent the degree at which I can consider C as a *subclass* of D.

Similarly, I disambiguate each occurrence of $w_{ar,i}$ in Arabic sentences with a word sense C_i and D_i in English and Arabic respectively. The distinct set of concepts $\bar{C}$ and $\bar{D}$ have been used to disambiguate the Arabic word w_{ar} respectively in senses from English and Arabic. For each $D_i \in \bar{D}$ I count the number that D_i has been co-disambiguated with every $C_i \in \bar{C}$. The proportion of the co-disambiguation for the two concepts D and C represent the confidence level at which I can consider D as a subclass of C. Then I use the F-measure to interpret the confidence level of the equivalent relation that aligns the concepts C and D.

For example, in Figure 4.1, in the sentence "the student seat around the *table* to eat their lunch", the words table and مائدة (pronounced Ma'ad) indicates the same meaning (a table at which meals are served). If most of the speakers would co-disambiguate "table" with the English word sense $Table_3^n$ [2], and with the Arabic word sense {طاولة Tawlah, منضدة Mndada, مائدة Ma'ad, سفرة Soufra}. Then $Table_3^n$ and {طاولة , منضدة , مائدة , سفرة} indicate the same concept.

Although I use a classification task that differs from the one proposed in [9], I can still use the inference rule they proposed to reason about mappings, to infer new mappings from existing mappings. Moreover, using the CL-WSD function as a

[2]The third noun sense in WordNet for table - a piece of furniture with tableware for a meal laid out on it

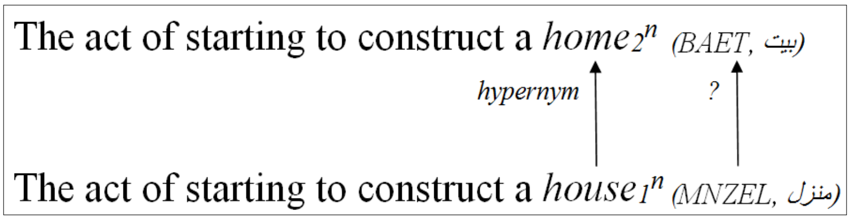


Figure 4.2: Example of CL-WSD task and a possible inference

classification task to evaluate the existence of relations among concepts, I can define a method to establish reference relationships between concepts by performing CL-WSD on sentence corpora.

For example, for each English word sense, a number of bilingual speakers (lexicographers) are asked to provide the equivalent Arabic word sense. For each word sense, the lexicographers substitute the English word with an Arabic concept. Using available bilingual dictionaries the lexicographers select the best translation. In Figure 4.2, in the sentence "the act of starting to construct a house", the English word "house" was CL-WSD with the English sense $House_n^1$ and the Arabic sense (منزل, Mnzel). For the same sentence I substitute the sense $House_n^1$ with its direct hypernym (subclass) sense $Home_n^1$ from the WordNet. If speakers CL-WSD the sense $Home_n^1$ with the Arabic sense (بيت, Baet). Ideally, I should be able to deduce the subclass (hypernym) relation between (منزل, Mnzel) and (بيت, Baet).

However, as mentioned before, not every concept is lexicalized in both (all) languages. The mappings, thus, obtained will form an initial semantic network. However, conflicts and overlaps might exist. Top levels concepts [61] can control and eliminate part of this problem. For example, the associated concepts should be classified under the same top concept. This direction of work also taking into account the relations confidence level will be pursued as a future work.

4.3.2 Challenges in the Experiment

It might be difficult and costly to conduct such experiment, as described above in Section 4.3.1, at a large scale. For instance, I need to collect feedback from large number of lexicographers. For a given language L_1, which has N^{L_1} word senses, M sentences are required in which K lexicographers are requested to perform CL-WSD tasks. That is, in order to obtain subclass relations between two languages, e.g., between English (L_1) and Arabic (L_2), the required number of CL-WSD tasks can be defined as follow: #CL-WSD tasks = N^{en}*M*K. In the same way, in order to obtain subclass relations between Arabic and English, the required number of CL-WSD tasks are given as follow: #CL-WSD tasks = N^{ar}*M*K.

Based on state-of-the-art well practices 3-10 sentences are required to provide a good quality for a cross-lingual word sense disambiguation tasks [86], and 3-5 lexicographers (users) are expected to validate each task [100]. For example, for mapping Arabic concepts to the English WordNet, which has more than 206 thousand word senses, at least about four Million (N^{en}*M*K + N^{ar}*M*K) CL-WSD tasks is required to establish equivalent relations between English and Arabic concepts.

One way to reduce the lexicographers' efforts and speed up the process, is to use available sense annotated corpora. Nevertheless, such an Arabic corpus is not available, which is the case for several resource-poor languages. Given this limitation, next I present two approaches to overcome this challenge.

Approach 1: using sense annotated corpora
I propose to mine the subclass relations starting from a sense annotated English corpus, I CL-WSD the English words with the equivalent Arabic senses, and then check if these relations can be converted to equivalence relations by exploiting the structure (relations) of the WordNet.

Figure 4.3 graphically depicts the experiment proposed to define cross-lingual ontology alignments, and in particular reference alignments, by assigning classification tasks to bilingual speakers. I propose to consider the Core WordNet concepts (5,000 concepts), which should represent shared concepts among different languages [19]. The majority of speakers is simulated by incorporation larger number of bilingual speakers (lexicographers). I suggest adopting a crowdsourcing model (e.g., use Amazon Mechanical Turkey [100]) to collect feedback from larger number of lexicographers. In order to improve the coverage and quality of translation, I plan to incorporate multilingual knowledge systems (e.g., Wikipedia, BabelNet [86]) to support the translation tasks. For example, by adopting cross-lingual explicit semantic analysis approach, CL-ESA, similar to previous works [23, 83].

For instance, the proposed experiment corresponds to a classification task; asking bilingual speakers to perform a $CL{-}WSD_{[En,Ar]}(w_i)$ tasks. I collect sentences from "Princeton Annotated Gloss Corpus" [19], a corpus of manually annotated WordNet synset definitions (glosses). Also, I collect sentences from SemCor [80], a

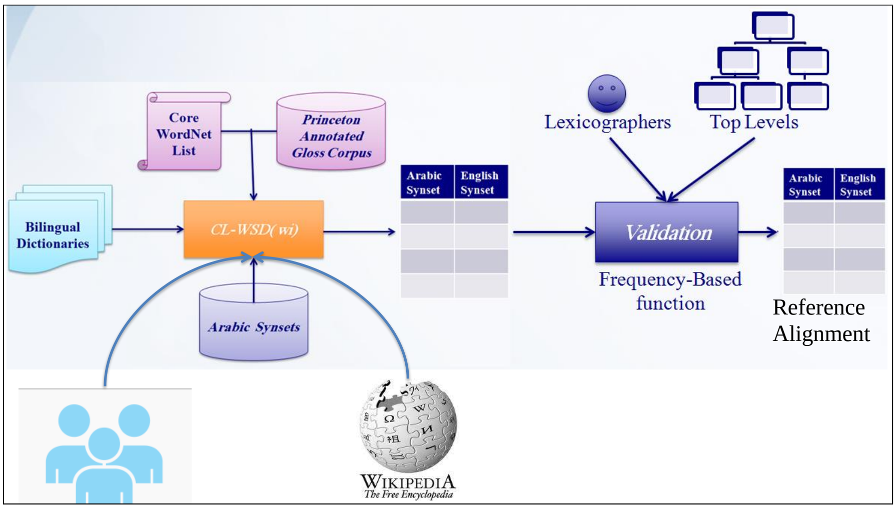


FIGURE 4.3: Using the cross-lingual mapping for building reference alignment

corpus of more than 200,000 words annotated with WordNet senses. The selected sentences are annotated with at least one sense that belongs to "Core WordNet". The reason for selecting Core WordNet concepts is that they represent the most frequent and salient concepts and thus can shared among many or most languages. Accordingly, I hypothesize that mapping the core WordNet concepts to the equivalent Arabic concepts will form the core for a lexical ontology for Arabic. Then I can extend it to include more cultural and language-specific concepts.

A significance result of a full-scale version of the proposed experiment is to generate a gold standard for cross-language mappings, which can be used to assess the various automatic cross-language matching systems as well to validate the proposed semantic mapping framework.

Approach 2 : using automatic translation

Using automatic translation tools is one option to reduce the number of user interaction and speed up the validation tasks. In BabelNet [86], they automatically translate a large set of English senses-tagged sentences using Google Translate, which are mostly extracted from Wikipedia pages. They translated (the mapped) WordNet and Wikipedia senses that occur at least in 3 different sentences, the sentences were collected from SemCor [80], and sentences from Wikipedia containing a (hypertext) link to the English Wikipedia page (concept). After they applied the automatic translation, the most frequent translation is detected and included as a variant for the English senses in the given language.

I have analyzed the quality and lexical richness of this approach (discussed in

Chapter 5), I noticed that this approach has a limited coverage and affects the mapping results. Based on these findings, I defined a lexical-based disambiguation algorithm which outperform this approach (discussed in Chapter 6). Moreover, I realized that validating mapping tasks can be performed at different level of efforts (discussed in Chapter 5), i.e., not a mere crowdsourcing model where tasks are equally processed. For instance, most, but not all, monosemous concepts can be mapped confidently even with simple selection methods (e.g., translation-based majority voting); whereas mappings involving polysemous but synonymless synsets are harder to filter out within the mapping selection task. In the first cases less number of lexicographers are required to perform the CL-WSD task, since I am confident that the automatic system (even when using baseline method) can provide a good match; whereas in the second case more attention is needed from a higher number of lexicographers to select the correct mappings. For example, the correct mapping is not always in the top-ranked matches, so lexicographers need to check more possible matches. Following these findings, I define an interactive cross-lingual mapping Web tool, which distributes the CL-WSD (mapping) tasks on users based on the estimation of the tasks difficulties, I discuss this approach in Chapter 6.

4.4 Conclusions

In this chapter, I discussed the semantics of cross-lingual mappings. I presented a classification-based semantic for cross-lingual ontology matching. An experimental setting is also introduced, through which the proposed mapping semantics can be evaluated and a gold standard dataset can be generated.

Because of the size of lexical resources, manual validation requires considerable effort, which makes it unfeasible at large scale. To overcome this limitation, automatic cross-lingual matching methods can be used either to compute mappings automatically, even at the price of accuracy, or to support semi-automatic mapping workflows by recommending mappings to lexicographers [93]. A natural subsequent step, is to investigate the extent to which the process of semi-automated creation is suitable for creating a lexical ontology.

For this goal, I plan to adopt crowdsourcing model [Sarasua et al. 2012] to collect users' feedback so as to converge on a set of shared agreed mappings. For such scenarios, i.e., establishing mappings between unstructured lexical resources (e.g., dictionaries, or collections of synsets) and structured resources like the English WordNet, mappings can be found using a mapping approach as the one described in this chapter. After the mappings are established, new relations between unstructured lexical elements can be derived from the relations between the concepts in the structured ontology, following the merge approach [115]. The experiment based on classification tasks assigned to bilingual speakers requires a large number of human inputs and can be practically difficult at large scale (e.g., for mapping dozens of thousands of concepts). Therefore, I plan to semi-automate the CL-WSD tasks: given a synset in one language the mapping algorithm is used to present to users a set of top-k English WordNet concepts matching the synset; the users classify the synset as equivalent to one of the suggested WordNet concept. Moreover, details on this approach are provided in Chapter 6. Through these experiments I wouldbe able also to validate the language-dependence hypothesis of the salient (core) concepts [19], especially when considering different families of languages.

Chapter 5

Effectiveness of Automatic Translations in Cross-Lingual Mapping Tasks: a Large Scale Study

5.1 Chapter Overview

Most of the approaches to map two ontologies lexicalized in different languages include a step in which the concepts' lexicalizations of one ontology are translated into the language of the other ontology. To translate concepts' lexicalizations, external *translation resources* such as machine translation tools or bilingual dictionaries have to be used. In this chapter I present a large-scale study on the effectiveness and quality of translations obtained by translation resources to match concepts lexicalized in different languages.

The study is organized as follows. By focusing on concepts' lexicalizations, I consider concepts as *synsets*, i.e., sets of words with an equivalent meaning in a given context [79]. This definition supports the idea of concept (synset) categorization based on varying characteristics, such as, word ambiguity (monosemous vs polysemous), number of synonyms (if any) a synset contains (e.g., synonymless vs synonymful), and concept position in a concept hierarchy (leaf node vs intermediate node). I define different measures to evaluate the translation effectiveness.

TABLE 5.1: Word distribution in the gold standards by category: quantity (percentage)

Words	English		Arabic		Italian		Slovene		Spanish	
Monosemous(M)	120433	(81.8)	10025	(72.3)	29816	(74.2)	28635	(71.6)	30106	(81.6)
Polysemous(P)	26873	(18.2)	3841	(27.7)	10362	(25.8)	11350	(28.4)	6774	(18.4)
Simple(S)	83118	(56.4)	8953	(64.6)	33133	(82.5)	29943	(74.9)	22630	(61.4)
Collection(C)	64188	(43.6)	4913	(35.4)	7045	(17.5)	10042	(25.1)	14250	(38.6)
M&S	59021	(40.1)	5361	(38.5)	22987	(57.2)	19223	(48.1)	16212	(44.0)
M&C	61412	(41.6)	4664	(33.6)	6827	(17.0)	9412	(23.5)	13894	(37.7)
P&S	24097	(16.4)	3592	(26.0)	10146	(25.3)	10720	(26.8)	6418	(17.4)
P&C	2776	(01.9)	249	(01.8)	218	(00.5)	630	(01.6)	356	(00.9)

Based on the afore-mentioned characteristics, I define different measures to evaluate translation effectiveness where I analyze the difficulty of the cross-lingual matching task w.r.t translation output. I first analyze the coverage of translations resources and its impact on the candidate match retrieval task. Further, I analyze the difficulty of the mapping selection task using a baseline mapping selection method. In this analyses, I use several measures based on a comparison with gold standards, i.e., mappings between concepts lexicalized in different languages that are deemed to be correct.

Next, in Section 5.2 I discus the concepts classification. Then, the evaluation measures used in the study are explained in Sections 5.3. In section 5.4, I present the experiments while conclusions end the chapter.

The scientific accomplishment presented in this chapter have been published in [1, 4, 5].

5.2 Classification of Concepts in Lexical Ontologies

In this section I discuses the lexical characteristics of concept in the gold standards (presented in Section 2.5.1).

Table 5.1 shows the distribution of words in each wordnet disaggregated by several categories: By considering *word ambiguity*, I distinguish between *Monosemous* words (M), words that have only one sense (meaning), and *Polysemous* words (P), words that have two or more senses. By considering *word complexity*, I distinguish between *Single* words (S), strings (lexemes) that have no spaces or hyphens, or

TABLE 5.2: Synsets categories

Category	Synset name	Definition *"synsets that have..."*
all_M	*all words Monosemous*	only monosemous words
all_P	*all words Polysemous*	only polysemous words
OWS	*One-Word*	only one word (*synonymless synset*)
MWS	*Many-Words*	two or more synonym words (*synonymful synset*)
M&OWS	*Monosemous and OWS*	only one word, which is also a monosemous word
M&MWS	*Monosemous and MWS*	two or more synonym words, which are all monosemous words
MIX	*MIXed*	monosemous and polysemous synonym words
P&OWS	*Polysemous and OWS*	only one word, which is also a polysemous word
P&MWS	*Polysemous and MWS*	two or more synonym words, which are polysemous words

TABLE 5.3: Synset examples for all categories in English

Category	Example	Definition
M&OWS	{desk}	a piece of furniture with a writing surface and usually drawers or other compartments
M&MWS	{tourism, touristry}	the business of providing services to tourists
MIX	{table$^+$, tabular array}	a set of data arranged in rows and columns
P&OWS	{cocktail$^+$}	a short mixed drink
P&OWS	{cocktail$^+$}	an appetizer served as a first course at a meal
P&OWS	{table$^+$}	a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs
P&MWS	{board$^+$, table$^+$}	food or meals in general

Collection words (C), strings that consist of two or more simple words, which are connected by spaces or hyphens. I also consider the four categories that are derived by combining word ambiguity and complexity categories. For example, *"tourism"* is a monosemous and simple word ($M\&S$), *"tabular array"* is a monosemous and collection word ($M\&C$), *"table$^+$"* is a polysemous and simple word ($P\&S$), and *"break up$^+$"* is a polysemous and collection word ($P\&C$).

Observation 1. A vast majority of collection words are monosemous words: only an average of 1.3% words are polysemous collection words across all wordnets. This means that a word used as concept label is less likely to be ambiguous if it is a composite word and more likely to be ambiguous if it is a simple word.

I classify the synsets based on the ambiguity and number of their words (respectively first and second, and third and fourth categories of synsets described in the upper part of Table 5.2). By combining these orthogonal classifications, I can consider five categories of synsets as described in the lower part of Table 5.2. One can observe that the *M&OWS* and the *M&MWS* are subsets of the *all_M*. The *P&OWS* and the *P&MWS* are subsets of the *all_P*, and the *MIX* are subsets of the *MWS*. Examples of synsets in the English WordNet for each category are shown in Table 5.3. Table 5.4 describes, for every wordnet, the total number and percentage of synsets grouped by category.

TABLE 5.4: Synset category-wise distribution in gold standards: quantity (percentage)

Synsets	English		Arabic		Italian		Slovene		Spanish	
all_M	57415	(48.8)	3381	(32.7)	14393	(42.7)	17615	(41.4)	19020	(49.1)
all_P	41568	(35.3)	4409	(42.6)	14641	(43.4)	19609	(46.0)	16269	(42.1)
MWS	53784	(45.7)	6162	(59.5)	13644	(40.4)	14994	(35.2)	14994	(38.7)
OWS	63875	(54.3)	4197	(40.5)	21084	(59.6)	27589	(64.8)	27589	(71.3)
M&OWS	33596	(28.6)	1995	(19.3)	10492	(31.1)	14848	(34.9)	14120	(36.5)
M&MWS	23819	(20.2)	1386	(13.4)	3901	(11.6)	2767	(06.5)	4900	(12.7)
MIX	18676	(15.9)	2559	(24.7)	5691	(16.9)	5359	(12.6)	3413	(08.8)
P&OWS	30279	(25.7)	2194	(21.2)	9609	(28.5)	12741	(29.9)	12005	(31.0)
P&MWS	11289	(9.60)	2215	(21.4)	4046	(12.0)	6868	(16.1)	4264	(11.0)

Observation 2. Wordnets have more synonymless synsets (OWS) than synonymful synsets (MWS), with 58.1% of synsets being, on average, synonymless. Arabic, which has less OWS than MWS, represents an exception among the considered wordnets. In particular, the Arabic polysemous synsets (all_P) are equally distributed between OWS and MWS.

In the gold standards (in Section 2.5.1) there exist mappings between synsets of every category. Examples of mappings for each couple of categories of synsets from Italian to English are shown in Table 5.5. The percentage of the mapped synsets between the non-English wordnets and the English WordNet, grouped by category, is reported in Table 5.6.

The results confirm that languages do not cover the same number of words as noticed in [52], and, hence, concepts shared in different languages have different ways to express their meanings (i.e., they belong to different lexical categories). For instance, 57% of the Italian $M\&OWS$ synsets are mapped to *monosemous* synsets in English ($M\&OWS$ and $M\&MWS$). On the other hand, 25% of the Italian $M\&OWS$ are mapped to *polysemous* synsets in English ($P\&OWS$ and $P\&MWS$). The percentage of monosemous non-English synsets that are mapped to the polysemous English synsets ranges from 10% (Slovene) to 30% (Arabic). The percentage of the monosemous English synsets that are mapped to the polysemous non-English synsets ranges from 6% (Arabic) to 14% (Italian). For instance, the $M\&OWS$ Italian synsets {fotografare} and {azioni ordinarie} are mapped to {shoot$^+$, snap$^+$, photograph$^+$} and {common shares, common stock, ordinary shares}, respectively a $P\&MWS$ and an$M\&MWS$ English synset.

Observation 3. Synsets in different languages, which have an equivalent meaning, can fall in different synset categories. For example, the Italian monosemous

TABLE 5.5: Examples of mappings between Italian and English synsets by category

Synsets	M&OWS		M&MWS		MIX		P&OWS		P&MWS	
M&OWS	{scuola d'arte}	{art school}	{radiostazione, stazione radio}	{radio station}	{tavolino, banco$^+$, scrivania}	{desk}	{ordinario$^+$}	{full professor}	{entita'$^+$, cosa$^+$}	{entity}
M&MWS	{turismo}	{tourism, touristry}	{accoppiata, abbinata}	{exacta, perfecta}	{docente$^+$, cattedratico, professore}	{prof, professor}	{viaggiatore$^+$}	{traveler, traveller}	{classe$^+$, aula$^+$}	{classroom, schoolroom}
MIX	{minorit}	{minority$^+$, nonage}	{biforcarsi, ramificarsi, diramarsi}	{branch, fork$^+$, furcate, ramify, separate}	{tavola$^+$, tabella}	{table$^+$, tabular array}	{contribuire$^+$}	{conduce, contribute, *lead*$^+$}	{cibo$^+$, pasto$^+$, mangiare$^+$}	{repast, meal$^+$}
P&OWS	{forchetta}	{fork$^+$}	{stretto, vicino}	{close$^+$}	{poltrona$^+$, seggiola, sedia}	{*chair*$^+$}	{*cosa*$^+$}	{*thing*$^+$}	{tavola$^+$, tavolo$^+$}	{table$^+$}
P&MWS	{chiudersi}	{close$^+$, shut$^+$}	{inquietarsi, allarmarsi}	{care$^+$, worry$^+$}	{segnare$^+$, scalfire}	{score$^+$, mark$^+$, nock$^+$}	{moderare$^+$}	{chair$^+$, moderate$^+$, lead$^+$}	{cibo$^+$, vitto$^+$}	{board$^+$, table$^+$}

TABLE 5.6: Distribution of mapping by category: percentage

English	M&OWS	M&MWS	MIX	P&OWS	P&MWS	M&OWS	M&MWS	MIX	P&OWS	P&MWS
	Arabic					**Italian**				
M&OWS	32.9	19.2	5.1	5.4	2.3	36.2	20.9	10.6	9.4	4.1
M&MWS	15.1	28.6	5.1	2.5	1.5	21.2	34.9	10.3	4.6	2.8
MIX	17.2	28.7	37.7	15.5	22.6	17.8	27.2	38.7	22.5	26.8
P&OWS	27.4	14.8	21.7	57.3	29.5	17.9	10.7	18.4	43.0	29.0
P&MWS	7.3	8.7	30.4	19.4	44.2	6.9	6.3	22.0	20.5	37.4
	Slovene					**Spanish**				
M&OWS	23.4	25.2	14.2	9.0	6.8	42.6	10.7	7.8	8.4	3.1
M&MWS	47.8	39.7	13.0	4.4	4.3	22.2	63.1	7.7	3.3	1.9
MIX	18.1	27.5	48.7	20.2	27.1	14.5	17.1	44.1	19.4	24.2
P&OWS	7.1	4.0	8.4	45.3	25.7	17.8	5.4	15.1	48.5	25.9
P&MWS	3.5	3.7	15.7	21.1	36.1	2.9	3.8	25.3	20.4	44.9

synonymless synset {forchetta} is mapped to the polysemous synomymless synset {fork$^+$} in English. This indicates that the monosemous word heuristic, which is adopted by some approaches to concept mapping and multilingual ontology construction, e.g., in [86], is successful for a large number of concepts but fails for still a relevant number of concepts. An average of 19.3% non-English monosemous synsets are mapped to English polysemous synsets in the gold standards, and an average of 8.9% English monosemous synsets are mapped to non-English polysemous synsets in the gold standards. More details on the impact of the monosemous word heuristics are provided in Section 5.4.1, where translation correctness is analyzed.

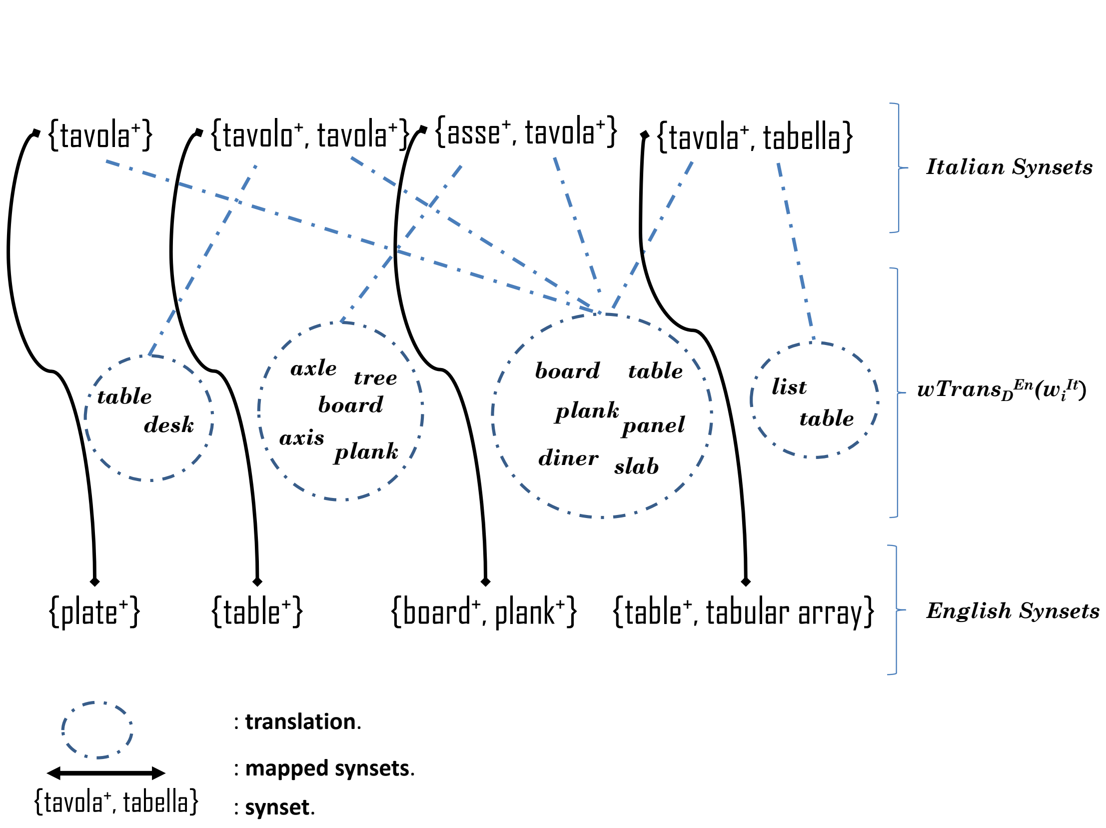


FIGURE 5.1: Example: Synset-translation

5.3 Evaluation Measures

In this study, I want to estimate the effectiveness of of translations obtained from multilingual lexical resources (hereby referred to as just *resources* in the rest of the paper) in finding candidate matches for a large set of concepts. I also want to estimate the difficulty of selecting one correct mapping among a set of candidate matches, based on the information provided by translations.

For the first objective, I define four measures that I use to evaluate translation correctness and coverage. The first two measures, **translation correctness** and **word sense coverage**, are used to evaluate the effectiveness of word-translations for a given word independently of its meaning, i.e., when the sense of the word is not given. The other two measures, **synset coverage** and **synonym coverage**, are used to evaluate the effectiveness of synset-translation for a given synset focusing on the lexicalization of the synsets in the target language. Word sense coverage and synset coverage are two measures proposed in previous work [86], but I rewrite their definitions according to the notation introduced in Section 2.4. Translation correctness and synonym coverage are introduced in this study. To facilitate the definition of these measures I first introduce the definition of *perfect translations* with respect to a gold standard. From these measures I can derive several measures, e.g., by averaging their values across one wordnet, to present the results of the experiments.

For the second objective, I use a measure that is straightforwardly derived from the well-known Precision measure [103] and is explained, directly, in Section 5.4.2.

5.3.1 Perfect Translations with a Gold Standard

The **perfect word-translation** of a word w^{L_1} into a target language L_2 w.r.t a gold standard gs is the set of every synonym words in all the possible senses of w^{L_1} in a target language L_2:

$$wTrans_{gs}^{L_2}(w^{L_1}) = \left\{ \bigcup_{i=1}^{n} w_i^{L_2} \mid \exists s^{L_2}(s^{L_2} \in senses_{gs}^{L_2}(w^{L_1}) \wedge w_i^{L_1} \in s^{L_2}) \right\} \tag{5.1}$$

Example 3. Figure 5.1 illustrates the synset-translation tasks for four Italian synsets into English. Each synset is mapped to its equivalent synset in English as specified by a gold standard gs. The translations are also obtained from the mappings between the Italian and the English wordnets represented in gs. For instance, the four (Italian $\leftrightarrow$ English) synsets mappings are: {tavola$^+$, tabella$^+$} $\leftrightarrow$ {table$^+$, tabular array}, {asse$^+$, tavola$^+$} $\leftrightarrow$ {board$^+$, plank$^+$}, {tavolo$^+$, tavola$^+$} $\leftrightarrow$ {table$^+$}, and {tavola$^+$} $\leftrightarrow$ {plate$^+$}. In Figure 5.1, the perfect word-translation of the Italian word *"tavola"* into English can be given as follow: $wTrans_{gs}^{En}(tavola^{+,It}) = \{table^+,\ tabular\,array,\ board^+,\ plank^+,\ plate^+\}^{En}$.

Observe that the perfect word-translation function returns every word of every possible sense in the target language, i.e., a word translation is perfect when it returns the complete lexicalization of every possible senses of an input word in the target language. This definition is motivated by the scope of this analysis, which evaluates the effectiveness of automatic translations in settings where the domain is not determined a-priori. When an individual input word is considered outside of a specific context, e.g., a specific sentence, a specialized domain or a concept hierarchy, the meaning of the word cannot be disambiguated, unless the word is monosemous. Otherwise, I observe that a domain-specific machine translation system, e.g., specialized in the financial domain, could determine the correct meaning (and translation) of a word, even when the word is considered individually, because of an implicit interpretation of the context by the system. Thus, in consideration of polysemous words and in absence of context specification, I defined a translation of a word (i.e., the set of words returned by a translation resource) *perfect* when it contains, for every possible usage of this word, all possible lexicalizations in the target language. If one considers word-translations in some specialized domain, he/she may need to adapt the definition of perfect word-translation consequently.

The **perfect synset-translation** of a synset s^{L_1} into a target language L_2 w.r.t a gold standard gs is defined as the set of every synonym words of the synset in L_2 mapped to s^{L_1} in gs. The perfect synset-translations can be defined as follows:

$$sTrans_{gs}^{L_2}(s^{L_1}) = \left\{ \bigcup_{i=1}^{n} w_i^{L_2} \mid \exists s^{L_2}(w_i^{L_2} \in s^{L_2} \wedge s^{L_1} \leftrightarrow s^{L_2}) \right\} \tag{5.2}$$

Example 4. In Figure 5.1, the perfect synset-translation of the Italian synset $\{$tavola$^+$, tabella$\}$ can be given as follow: $sTrans_{gs}^{En}(\{tavola^+, \, tabella\}^{It}) = \{table^+, \, tabular\ array\}^{En}$.

5.3.2 Evaluation of Word-Translations

Translation correctness. For an input word, this measure evaluates to what extent a resource returns precise and complete translations when compared to perfect word translations defined by a gold standard, which consider every possible sense of the word in the target language.

To define this measure, I need to specify when a word returned by a resource is correct. A word w^{L_2} is a **correct-translation** for a word w^{L_1} w.r.t a gold standard gs, if w^{L_2} belongs to the set of perfect word-translations for w^{L_1} w.r.t gs (denoted by $wTrans_{gs}^{L_2}(w^{L_1})$). This principle is captured by the function $correctT_{w^{L_1},D}(w^{L_2})$ defined by the following equation:

$$correctT_{w^{L_1},D}(w^{L_2}) = \begin{cases} 1 & \text{if } w^{L_2} \in \{wTrans_{gs}^{L_2}(w^{L_1})\}. \\ 0 & \text{otherwise.} \end{cases}$$

Example 5. In Figure 5.1 the English words *"table", "board",* and *"plank"* are correct translations for the Italian word *"tavola"*, e.g., $correctT_{tavoal^{It}}$ *(*$table^{En}$*) =1.* The English words *"diner", "panel",* and *"slab"* are *incorrect* translations for the Italian word *"tavola"*, e.g., $correctT_{tavoal^{It}}(slab^{En})=0$.

I measure the correctness of translations returned by a resource D for a word w^{L_1} with *translation-correctness* as defined in Eq. 5.3. The measure is computed as the harmonic mean, i.e., F_1*-measure*, of two measures: 1) *Precision (Pr)*, defined as the number of correct translations returned by the translation resource D over the total number of translations returned by D; 2) *Recall (R)*[1], defined as the number of correct translations returned by the translation resource D over the total number of perfect word translations. I use Recall, Precision and F_1*-measure* (computed with its standard range), but normalized in the range [0..100]. When no translation is returned by the resource D, Precision is set to zero.

$$Pr = \frac{|\{w^{L_2}|correctT_{w^{L_1},D}(w^{L_2})\}|}{|\{wTrans_D^{L_2}(w^{L_1})\}|} * 100, R = \frac{|\{w^{L_2}|correctT_{w^{L_1},D}(w^{L_2})\}|}{|\{wTrans_{gs}^{L_2}(w^{L_1})\}|} * 100$$

$$TransCorrectness_D^{L_2}(w^{L_1}) = F_1(Pr, R) * 100 = 2\frac{Pr * R}{Pr + R} * 100 \tag{5.3}$$

Example 6. In the example shown in Figure 5.1, the correctness of English translation of the Italian word *"tavola"* is computed as follows: recall $R = 60.0$, precision $Pr = 50.0$, and the translation-correctness $TransCorrectness^{En}(tavola^{It}) = 55.0$.

Word Sense Coverage. For an input word, this measure evaluates how many of its possible word senses in a target language are covered at least by a word translation (as defined in [86]). A translation *covers a sense* s^{L_2} of an input word w^{L_1} in a different language when the translation resource returns at least one word of s^{L_2}. I use the binary predicate $cov(x, y)$ to state that a word-translation x covers the sense y. Word senses coverage tells to what extent the polisemy of a word is covered by a translation resource. Ideally, a resource is effective in translating a word w^{L_1} when it is able to return some correct-translations for every possible sense of w^{L_1} in L_2.

[1] I remark that Recall is also named *translation accuracy* in the WSD literature [85].

Given a word w^{L_1} translated into a target language L_2 with a resource D, the word senses coverage of w^{L_1} is defined as follows:

$$wsCoverage_D^{L_2}(w^{L_1}) = \frac{|\{s^{L_2} \mid s^{L_2} \in senses_{gs}^{L_2}(w^{L_1}) \wedge cov(wTrans_D^{L_2}(w^{L_1}), s^{L_2})\}|}{|\{s^{L_2} \mid s^{L_2} \in senses_{gs}^{L_2}(w^{L_1})\}|} \tag{5.4}$$

Example 7. In Figure 5.1, the polysemous Italian word *"tavola"* has four senses, each one is mapped into an equivalent synset in English. Using the translation resource D three out of the four senses are covered (Eq.5.4). For instance, the senses {table$^+$} is covered, $cov(wTrans_D^{En}\ (tavola), \{table\}) = 1$, while the sense {plate$^+$} is *not* covered, $cov(wTrans_D^{En}(tavola), \{plate\}) = 0$.

5.3.3 Evaluation of Synset-Translations

Synset coverage. This measure is defined as boolean function applied to an input synset. A synset s^{L_1} is covered if its translation, i.e., the multi set union of the translation of its constituent words, returns at least one word of its equivalent synset in the target language. This measure is useful when computed for a set of source synsets as in previous work [86]. For example, by computing the percentage of source synsets mapped in a gold standard that are covered by a translation resource, I can evaluate the number of mappings that can be discovered by using this translation resource.

To formally define synset coverage in a compact way, I can use the concept of perfect synset translation for a synset s^{L_1} in a target language L^{L_2}, denoted by $sTrans_{gs}^{L_2}(s^{L_1})$. If s^{L_1} is a synset translated in a target language L_2 with a resource D, synset coverage is defined as follows:

$$sCoverage_D^{L_2}(s^{L_1}) = \begin{cases} 1 & \text{if } \exists w^{L_2}(w^{L_2} \in sTrans_D^{L_2}(s^{L_1}) \wedge w^{L_2} \in sTrans_{gs}^{L_2}(s^{L_1}) \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

Example 8. Consider the Italian and their equivalent English synsets depicted in Figure 5.1. Three out of four Italian synsets are covered because their translation returns at least on word of their equivalent English synsets. For instance, the mapping {tavolo$^+$, tavola$^+$} $\leftrightarrow$ {table$^+$} is covered, while the mapping {tavola$^+$}

$\leftrightarrow$ {plate$^+$} is not covered.

Synonyms coverage. For an input synset s^{L_1}, this measure evaluates the number of words of s^{L_1} for which a word-translation covers the equivalent synset in the target language. This measure tells how many synonyms in a concept lexicalization are covered by correct translations.

Give a synset s^{L_1} having its equivalent synset s^{L_2} in the target language; s^{L_1} is translated using a resource D, then the synonym coverage of s^{L_1} is defined as follows:

$$synonymsCoverage_D^{L_2}(s^{L_1}) = \frac{|\{w^{L_1} \mid w^{L_1} \in s^{L_1} \wedge cov(wTrans_D^{L_2}(w^{L_1}), s^{L_2})\}|}{|s^{L_1}|} \tag{5.6}$$

Example 9. In Figure 5.1 the Italian synsets {tavola$^+$, tabella}, {asse$^+$, tavola$^+$}, and {tavolo$^+$, tavola$^+$} have full synonym words coverage (Eq.5.6). Whereas, the synset {tavola$^+$} is not covered because its only word is not covered.

Synonym coverage is a valuable measure to evaluate translation resources in the field of cross-lingual concept mapping. Consider, for example, an input synset s^{L_1} and a translation resource that returns many of the (synonym) words of its equivalent synsets s^{L_2} in the target language. On the one hand, these synonym words are useful to increase the probability of finding s^{L_2} among the candidate matches of s^{L_1}. On the other hand, these synonym words can be used as evidence for selecting s^{L_2} as the best match for s^{L_1}, e.g., if compared to other candidate matches for which little evidence is collected via translation[2]. Finally, I observe that synonym words coverage is a complementary indication of the word senses coverage to measure the effectives of a translation resource, i.e., the coverage of the synonym words is a tool to disambiguate the polisemy of translations returned by a translation resource.

[2]This intuition has been used, for example, in a cross-lingual similarity measure proposed to support matching of lexical ontologies lexicalized in different languages [3].

Throughout this paper, in order to quantify the overall coverage measures and correctness of the word-translation tasks across each dataset (wordnet), I compute the *Macro-average F-measure* [114]. The reported coverage measures are normalized in the range [0..100].

5.4 Experiments

Three experiments are conducted to study the coverage, correctness, and impact of automatic and external translation resources on mapping concepts lexicalized in different languages. Next, In Section 5.4.1 I evaluate the coverage and correctness of translations obtained with different resources to discuss their impact on retrieving candidate matches in concept mapping tasks. In Section 5.4.2, evidence collected from automatic translations is used in a baseline mapping selection approach, i.e., majority voting, to evaluate the difficulty of the mapping selection task. In Section 5.4.3, I analyze the coverage of translation resources in relation to the position of the concepts in the semantic hierarchies. Finally, in Section 5.4.4 I summarize the study observations and draw some potential future directions.

Experimental setup: The English, Arabic, Italian, Slovene, and Spanish wordnets (detailed in Section 2.5.1) are imported into a database. The wordnets database includes the words, synsets, semantic relations, and the mappings between each non-English wordnet and the English Wordnet. Then, I compiled different bilingual dictionaries (Table 2.1) with Google Translate API and BabelNet as described in Section 2.4.1. I stored the dictionaries in a database so as to efficiently execute the experiments.

Next, I describe in details the three experiments presented in this study.

5.4.1 Experiment 1: Coverage and Correctness of Translations for Candidate Matches Retrieval

In order to evaluate the **coverage** of different translation resources (i.e., coverage of the translations obtained with different resources) two measures are used. I compute the *average word sense coverage* across all words of a wordnet, where word sense coverage is defined for an individual word as in Eq.5.4. I compute

TABLE 5.7: Word sense and synset coverages with different translation settings

Translation	**Arabic**		**Italian**		**Slovene**		**Spanish**	
	Senses	Synsets	Senses	Synsets	Senses	Synsets	Senses	Synsets
BN_{core}	19.9	37.4	40.0	62.5	28.8	44.2	33.9	44.7
BN	30.8	51.3	51.7	72.8	35.9	52.0	39.8	49.0
MT_fromEn	51.3	69.9	60.2	81.9	40.2	60.0	56.1	67.8
MT_toEn	57.9	76.1	65.4	83.9	49.6	67.2	67.0	77.0
MT	59.2	77.7	68.1	87.6	53.8	72.4	69.4	79.7
$MT\&BN_{core}$	60.8	79.2	69.8	89.0	55.8	74.2	71.5	81.3
$MT\&BN$	62.5	80.2	72.2	89.9	57.5	75.2	72.3	81.7

the *average synset coverage* across all synsets of a wordnet, where synset coverage is (a boolean value) defined for an individual word as in Eq.5.5. All values are normalized in the range [0..100]. For sake of clarity I will simply refer to these measures as word sense and synset coverage (at the wordnet level).

Table 5.7 reports, for each wordnet, word sense and synset coverage with different translation settings. Synsets have higher coverage than word senses in all the translation settings. This can be explained with the observation that a synset is covered if its translation returns at least one word of the lexicalization of its equivalent synset in the target language (see Eq.5.5).

I observe that machine translations from non-English *to English* (MT_toEn) achieve higher word sense and synset coverage than machine translation *from English* to non-English (MT_fromEn). For instance, word sense coverage of MT_toEn is from 5.2 (Italian) to 10.9 (Spanish) percentage points higher than MT_fromEn, and synset coverage of MT_toEn is from 2.0 (Italian) to 9.2 (Spanish) percentage points higher than MT_fromEn.

Observation 4. Machine translation resources perform asymmetrically: MT_toEn achieves higher word sense and synset coverage than MT_fromEn.

The machine translation bilingual dictionary (MT), which I built from the union of both machine translation directions (see Section 2.4.1), performs better than the dictionaries that I built considering each direction alone (i.e., MT_fromEn or MT_toEn). Word sense coverage of MT is on average 2.7 and 8.2 percentage points higher than MT_toEn and MT_fromEn, respectively. Synset coverage of MT is on average 3.5 and 7.0 percentage points higher than MT_toEn and MT_fromEn, respectively.

BN_{core} and BN translation settings, which are based on BabelNet, obtain lower coverage than every machine translation setting for all wordnets. This can be explained by limited coverage of the words that occur in non-English wordnets by

Table 5.8: Average synset coverage by category

Synsets	Arabic			Italian			Slovene			Spanish		
	BN	MT	MT&BN	BN	MT	MT&BN	BN	MT	MT&BN	BN	MT	MT&BN
all_M	35.4	64.2	67.3	68.6	86.0	88.4	59.8	78.8	80.8	34.4	78.1	79.5
all_P	58.8	82.9	85.4	69.0	80.5	83.0	44.8	65.5	69.1	62.5	80.2	83.0
OWS	44.5	67.0	70.5	64.1	79.5	82.2	52.8	70.1	73.0	44.1	75.8	78.0
MWS	56.1	85.0	86.8	81.0	93.6	95.2	50.7	76.6	79.1	59.0	87.7	89.3
M&OWS	32.2	58.0	61.5	63.8	83.5	86.1	60.8	78.3	80.3	32.4	74.7	76.2
M&MWS	40.3	73.4	76.0	81.6	92.6	94.5	54.3	81.4	83.2	40.1	88.0	89.0
MIX	59.8	86.9	88.6	80.7	94.3	95.8	53.1	76.5	79.0	65.5	85.8	87.6
P&OWS	55.8	75.4	79.0	71.0	83.3	86.4	43.4	60.5	64.5	57.9	77.1	80.2
P&MWS	61.9	90.5	92.0	80.8	93.7	95.1	47.3	74.7	77.6	75.5	88.8	90.9

Wikipedia concepts (which mostly cover named entities), and by incompleteness of the mappings used to construct BabelNet [86]. However, it should be remarked that, for several languages, BabelNet also includes the lexicalizations from the Open Multilingual WordNet that have been excluded in this study because it is part of the gold standard (see Section 5.2). This means that for several well-known languages such as French, Germany, Spanish, and Italian[3] I can expect much higher translation coverage from BabelNet. Still, best results are obtained when combining all available translations, i.e., both from the machine translation tool and BabelNet, *MT&BN*. For instance, *MT&BN* word sense coverage is on average 3.5 percentage points higher than *MT*. *MT&BN* synset coverage is on average 2.4 percentage points higher than *MT*.

I also observe that *BN* achieves considerably higher coverage than BN_{core}, with an average difference in word sense and synset coverage of 10.4 and 10.1 percentage points respectively (BN_{core} is a subset of *BN* - see Section 2.4.1). However, most of this additional coverage is lost when combining BN_{core} and *BN* with *MT* translations: *MT&BN* word sense coverage is on average only 1.7 percentage points higher than $MT\&BN_{core}$, and *MT&BN* synset coverage is on average 0.8 percentage points higher than $MT\&BN_{core}$.

Observation 5. The results highlight that machine translation tools achieve higher coverage than BabelNet, which integrates several Web-based linguistic resources (i.e., Wikitionary, OmegaWiki, WikiData, and Wikipeida redirection links). However, integrating BabelNet with machine translation tools still yields a significant gain in coverage, mostly because of BN_{core} (Wikipedia inter-lingual links, and the context based translations).

Table 5.8 reports the average coverage for each synset category by using *BN, MT* and *MT&BN* translation settings (the settings achieving highest coverage).

[3]See the Languages and Coverage Statistics at http://babelnet.org/stats .

TABLE 5.9: Average number of candidate matches

Synsets	Arabic	Italian	Slovene	Spanish
synonymless (OWS)	48	17	11	27
synonymful (MWS)	124	49	21	75

The results show that the *synonymful* synsets (MWS) are covered more than *synonymless* synsets (OWS) for every wordnet and almost every translation setting. This confirms the intuition that richer concept lexicalizations help to find at least one correct translation using machine translation resources. *Polysemous* synsets (all_P) are covered more than the *monosemous* synsets (all_M) for Arabic and Spanish, but less than *monosemous synsets* (all_M) for Italian and Slovene. This can be explained by the distribution of polysemous and monosemous synsets between synonymless and synonymful synsets: most of the monosemous synsets (all_M) are synonymless synsets, and most of the polysemous synsets (all_P) are synonymful synsets. MIXed synsets are the most covered synsets, since they are synonymful synsets, which combine monosemous and polysemous words.

Observation 6. Synonymful synsets (MWS) are covered more than synonymless synsets (OWS) (see Table 5.8). However, a higher coverage comes at the price of a larger number of candidate matches, thus making the mapping selection task more challenging (see Table 5.9).

Observation 6 can be supported by figures shown in Table 5.9, which reports the average number of candidate matches for synonymless vs synonymful synsets. In addition, most of synonymful synsets contain at least one polysemous word (see Table 5.4). Thus, one can expect that the sets of candidate matches returned by translations of synonymful synsets are not only larger in size, but also noisier, because of the translation of polysemous words. A more in-depth analysis on the difficulty of the mapping selection task for the different synset categories is provided in Section 5.4.2. Such analysis will confirm that the mapping selection problem is more difficult for synsets that contain polysemous words, which represent the majority of synonymful synsets. At the same time, the joint translation of synonym words can support mapping selection for many synsets (e.g., for synsets that do not contain only polysemous words, e.g., $MIXed$ synsets), as a means to collect evidence for deciding about a mapping.

TABLE 5.10: Average recall and word-translation correctness by category

Words	Arabic						Italian					
	BN		MT		MT&BN		BN		MT		MT&BN	
M	20.2	(63.6)	45.1	(36.9)	48.0	(56.4)	49.0	(65.6)	65.8	(47.0)	69.8	(62.1)
P	53.1	(38.0)	83.3	(22.8)	85.2	(40.9)	71.5	(44.9)	89.8	(31.9)	91.3	(45.1)
S	38.1	(48.3)	67.0	(27.3)	70.0	(49.1)	54.4	(57.0)	73.0	(41.0)	75.9	(55.5)
C	13.3	(63.4)	35.1	(43.9)	37.0	(53.8)	56.9	(65.8)	67.3	(47.5)	72.6	(62.8)
M&S	26.7	(63.0)	54.8	(32.6)	58.6	(57.2)	46.7	(65.3)	65.4	(46.7)	69.1	(61.6)
M&C	12.7	(65.2)	34.0	(44.9)	35.8	(55.0)	56.9	(66.3)	66.9	(47.9)	72.2	(63.5)
P&S	55.0	(37.7)	85.2	(22.3)	87.0	(40.9)	71.8	(44.8)	90.0	(31.7)	91.4	(45.1)
P&C	25.3	(46.6)	55.4	(32.0)	58.2	(40.3)	56.0	(48.6)	80.3	(39.0)	84.9	(44.1)
ALL	29.3	(50.8)	55.7	(31.0)	58.3	(50.2)	54.8	(58.6)	72.0	(42.1)	75.3	(56.8)

Words	Slovene						Spanish					
	BN		MT		MT&BN		BN		MT		MT&BN	
M	45.2	(66.1)	63.6	(47.8)	66.6	(60.8)	28.1	(61.6)	68.4	(48.1)	70.8	(56.9)
P	42.6	(39.6)	73.0	(30.1)	75.4	(33.7)	74.8	(38.9)	92.1	(28.4)	93.7	(41.4)
S	43.7	(56.3)	67.0	(41.4)	69.7	(49.9)	48.9	(51.0)	78.8	(41.0)	81.0	(53.3)
C	46.7	(66.0)	64.3	(45.0)	67.3	(60.3)	17.4	(62.3)	63.1	(48.4)	65.7	(53.5)
M&S	43.7	(65.8)	62.5	(49.0)	65.4	(60.7)	37.6	(61.1)	73.2	(47.6)	75.5	(59.1)
M&C	48.3	(66.5)	66.0	(45.5)	69.1	(61.1)	17.0	(63.0)	62.8	(48.6)	65.4	(53.9)
P&S	43.7	(39.2)	75.0	(30.0)	77.4	(33.5)	77.3	(38.7)	93.0	(27.9)	94.7	(41.5)
P&C	22.9	(50.3)	38.9	(31.3)	41.1	(39.9)	30.6	(47.5)	75.6	(38.8)	76.7	(40.3)
ALL	44.5	(58.9)	66.3	(42.3)	69.1	(52.4)	36.7	(53.1)	72.8	(43.5)	75.0	(53.3)

In order to evaluate the **correctness** of different translation resources (i.e., correctness of the translations obtained with different resources) two measures are used. I compute the *average word-translation correctness* across all words of a wordnet; word-translation correctness is defined for an individual word as in Eq.5.3. In addition, I report *average word-translation recall* (*recall*, for short), using the subformula in Eq.5.3[4].

Average recall and word-translation correctness for the *BN*, *MT* and *MT&BN* dictionaries, disaggregated by word category, are reported in Table 5.10.

The results show that word-translation is more correct for monosemous and collection words than for polysemous and simple words. In contrast, recall of word-translation is higher for *polysemous* words (P) than for *monosemous* words (M) with every translation resources and for every wordnets, with the exception of the *BN* dictionary for the Slovene wordnet. Recall of word-translation is also higher for *simple* words (S) than for *collection* words (C) in every setting. These observations can be explained by monosemous words being usually less frequent and more domain-specific than polysemous words. In addition, most of collection words are also monosemous words - as remarked in *Observation 1* -, while most of polysemous words are simple words: recall and correctness of translations for simple words is affected by the translation of simple polysemous words. Translations

[4] Word-translation correctness is defined using a formula based on F_1-*measure*.

of polysemous and simple words return on average larger word sets. These sets are more likely to contain richer lexicalizations in the target language, but also to contain words that do not belong to any sense of the input words in the target language.

Observation 7. The translation of monosemous and collection words is on average more correct than the translation of polysemous and simple words, but achieves lower recall.

Focusing on the performance of different translation resources, I can notice that recall for MT is higher than for BN, while correctness for BN is higher than for MT. $MT\&BN$ combines the strengths of both dictionaries, i.e., higher recall of word-translation because of MT, and higher correctness because of BN. For instance, in Table 5.10 I can notice that the correctness of word-translation is improved by 9.8 and 19.2 percentage points for Spanish and Arabic respectively, if I add to MT translations derived from BN. The recall of word-translation is improved by as much as 20.4 and 38.3 percentage points for Italian and Spanish respectively, if I add to BN translations derived from MT. The best results are thus obtained for $MT\&BN$, for which I obtain recall (correctness) scores that range from 58.3%(50.2%) (Arabic) to 75.3% (56.8%)(Italian). The low recall for Arabic can be explained by a low recall for translations of monosemous collection words.

Observation 8. The combination of machine translation tools with Web-based linguistic resources and context-based sentence translations, like the ones incorporated in BabelNet, improves not only the recall, but also the correctness of word-translations.

5.4.2 Experiment 2: Mapping Selection Difficulty

On one hand, the translations returned for a given synset can be used as evidence to select a mapping to a synset in a target language. On the other hand, translations of many, polysemous words in a synset can return several candidate matches, most of which are incorrect, thus making the mapping selection task difficult to solve. This experiment analyzes the difficulty of the mapping selection task when performed over candidate matches retrieved with translation resources.

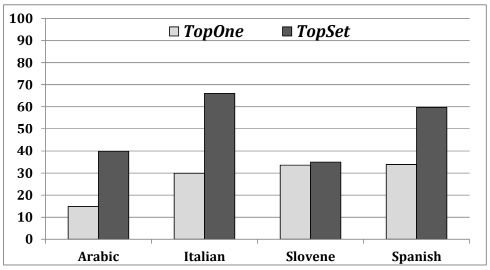
In this experiment I use translations returned by the MT machine translation tool, for the sake of simplicity (with the exception of the analysis of the synonym word coverage, where I also include BN). I focus on MT because, as shown in the previous section, it has higher coverage than BN, and it has been widely used in previous work on ontology matching. In addition, the slight increase in coverage obtained with $MT\&BN$, when compared to MT, can be ignored for this particular experiment.

To perform this analysis I use a greedy baseline method for candidate mapping selection and I compare the quality of the alignment computed with this method to the gold standard alignment. As a baseline mapping selection method I use *majority voting* on top of the evidence collected from synset translations.

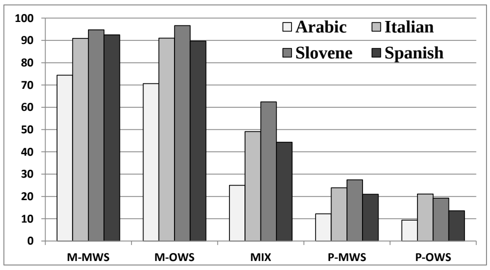
Mapping Selection with Majority Voting. Every source synset is translated using the synset translation function defined in Eq.2.1. The output is represented as the multi set union of the returned translations. Each word $w^{(i)}$ in the multi set, with (i) being the word frequency count, represents i votes for all the candidate matches that contain w. Therefore, a candidate match t for a source synset s, such that t contains many words returned by the translation of s, will receive more votes and will be more likely to be the target of the selected mapping. Candidate matches are ranked by votes and the mapping containing the top-voted match is selected.

It can happen that several candidate matches receive an equal number of votes, which results in a *tie*. In this case, for a source synset the mapping selection task is *undecidable*; in contrast I will say that a mapping is *decidable* when a unique candidate match receives the highest number of votes. However, when a tie occurs among a set of top-voted candidate matches, it is valuable to know if this set contains also the correct mapping (w.r.t the gold standard) and the number of candidate matches in the tie. In fact, if the set of top-voted candidate matches also contains the correct match, the correct mapping could be found via user interaction with relatively low effort. For these reasons I use two settings in the experiments with a majority voting candidate selection approach:

- **TopOne**: if there exists a unique top-voted candidate match for a source synset, the mapping containing this match is selected and included in the alignment. If a tie occurs, no mapping is selected.



(a) correct mappings found with *TopOne* and *TopSet* settings



(b) correct mappings in the *distinguishable* candidate matches by category

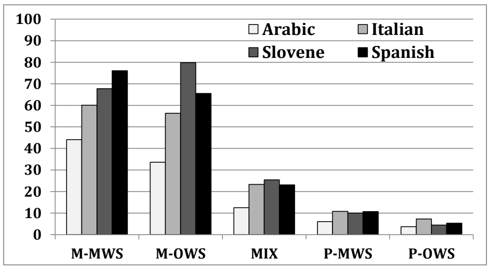
FIGURE 5.2: Correct mappings found with baseline selection strategy

- **TopSet**: the correct mapping is selected by an oracle from the set of top-voted matches (no matter of its cardinality) and included in the alignment.

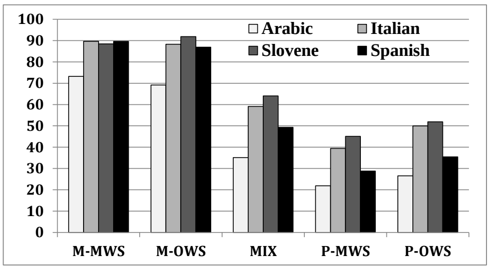
To quantify the quality of the alignment I compute *(selection) correctness* as the percentage of the correct mappings returned by each selection setting *over the set of covered mappings*, i.e., the mappings for which the set of candidate matches contains the correct mapping[5]. In other words, in *TopOne* setting, a mapping is considered correct for a source synset, only when the correct match for the synset (according to the gold standard) is its unique top-voted candidate match; in *TopSet* setting, a mapping is considered correct for a source synset, whenever the correct match for the synset is included in the set of its top-voted candidate matches. Observe that every mapping that is counted as correct in *TopOne* setting, is also counted as correct in *TopSet* setting.

A comparison between the performance in terms of correct mappings returned in the *TopOne* and *TopSet* selection settings for each wordnet is shown in Figure 5.2(a). The average of correct mappings obtained in *TopOne* and *TopSet* settings is 28% and 50% respectively. Based on the performance of such simple baseline methods, I suggest that translations can be helpful for mapping selection, although more sophisticated methods to make use of their evidence have to be devised. In addition, number of correct mappings can be increased up to an average of 30 points in the case where I assume that a user can select the correct mapping among the set of top-voted matches returned by a mapping selection method, e.g., with an interactive ontology matching approach [26, 100]. However, the average

[5]This is equivalent to compute a *relative precision* measure: Precision is interpreted as usual in ontology matching [103] but normalized in the range [0..100], and evaluated only over a restricted subset of the gold standard. Such a restricted subset consists of all the mappings containing source concepts that are covered by translations



(a)



(b)

FIGURE 5.3: Percentage of the correct mappings by synset category with (a) *TopOne* selection and (b) *TopSet* selection

cardinality of the sets of top-voted matches (*TopSet*) is as high as 49 synsets, which makes it difficult for users to make a decision.

Figure 5.2(b) shows, for every wordnet and every category of source synset, the percentage of correct mappings found using *TopOne* selection over the total of synset with decidable mappings. The baseline *TopOne* mapping selection strategy achieves a remarkable performance for monosemous synsets (i.e., *M&OWS* and *M&MWS*) and poor performance for polysemous synsets. On average, *TopOne* selection is capable to select correct matches for as much as 88.2% of the monosemous synsets.

Figure 5.3(a) and 5.3(b) show, for every wordnet and every category of target synset, the percentage of correct mappings found respectively with *TopOne* and *TopSet* selection settings. I figured out that mappings to synsets with polysemous words, in particular to *polysemous synonymless synsets* (*P&OWS*), are much more likely to be undecidable, i.e., a set of many top-voted candidate matches is found. In fact, when the target synsets are *P&OWS*, the mapping is almost always undecidable with the *TopOne* selection.

Observation 8. Evidence provided by machine translation tools is valuable to successfully decide upon correct mappings for monosemous synsets, while it fails to support such a decision for most of the polysemous synsets.

Observation 9. Mappings with polysemous and synonymless target synsets (*P&OWS*) cannot be successfully selected by leveraging only the evidence from translations and a simple selection strategy like majority voting because translations assign an equal number of votes to several candidate matches.

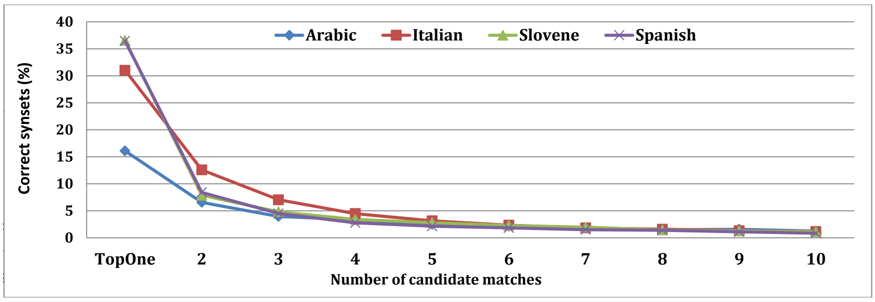


FIGURE 5.4: Percentage of correct mappings vs. size of top-voted candidate matches with *TopSet* selection

Observation 10. If the set of top-voted candidate matches can be validated, e.g., as in the *TopSet* selection settings, it is possible to find a correct mapping for a vast majority of monosemous synsets (on average, 85%).

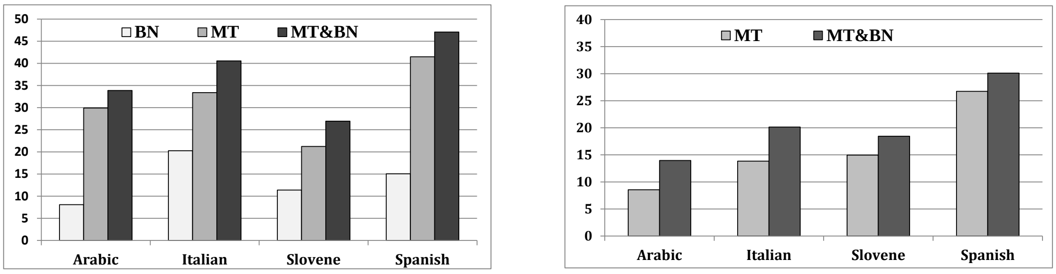
I want to investigate if correct mappings are more likely to be found when a larger or a small number of top-voted mappings is selected (with *TopSet* selection). To this end, I analyze the distribution of the correct mappings found with *TopSet* selection among top-voted candidate matches of different size for every wordnet. Correct mappings are found in sets of top-voted candidate matches with a size that ranges from 1 to 238 candidates. The distribution is plotted in Figure 5.4: *x-axis* represents the number of selected top-voted candidate matches (up to size equal to ten), while the *y-axis* represents the percentage of found correct mappings. On average, 28% of correct mappings are found when a unique top-voted candidate match exists, i.e., like in *TopOne* selection settings (see Figure 5.2(a)). For instance, about 4% of the correct mappings are found in sets of top-voted mappings that contain four candidate matches, a percentage that represents an absolute number of 317, 1455, 991, and 1328 synsets for the Arabic, Italian, Slovene, and Spanish wordnets, respectively.

Observation 11. Synsets that occur as targets in mappings found with *TopOne* selection (decidable mappings) can be safely filtered out from candidate matches for other source synsets, with an error estimated to be as low as 0.2% of removing a correct match.

Finally, I analyze the impact of synonyms on the mapping selection task. Synonymful synsets (i.e., *M&MWS*, *MIX*, and *P&MWS*) are more likely to be correctly mapped with *TopOne* selection (Figure 5.3(a)) than synonymless synsets (i.e., *M&OWS* and *P&OWS*), even if the average number of candidate matches is greater for synonymful synsets than for synonymless synsets (see Table 5.9).

TABLE 5.11: Synonym words coverage (%) for synonymful synsets (MWS)

Translation	Arabic	Italian	Slovene	Spanish
BN	51.9	59.8	56.7	61.2
MT	68.9	68.5	61.5	74.4
MT&BN	71.3	72.6	65.4	77.5



(a) Percentage of *MWS* synsets that fully covered with different translation settings

(b) Percentage of correct mappings for *MWS* synsets found with *TopOne* selection

FIGURE 5.5: Synonymful synsets (*MWS*) whose synonym words are fully covered

These results confirm that synonyms are helpful not only for retrieving candidate matches - as previously observed in *Observation 6* - but also for selecting the correct mappings: the translation of different words that express a same concept provide evidence to decide the best mapping for this concept.

Table 5.11 reports, for every wordnet, the synonym words coverage for synonymful synsets (*MWS*) using the *BN*, *MT* and *MT&BN* dictionaries (synonym words coverage is defined by Eq.5.6). The best results are obtained with *MT&BN*, with synonym words coverage ranging from 65.4% (Slovene) to 77.5% (Spanish). Thus, on average, more than two synonyms are translated correctly in synonymful synsets.

Figure 5.5(a) shows the percentage of synonymful synsets that are **fully covered**, i.e., synsets that contain only words that are correctly translated. On average, the *MT* dictionary fully covers a greater percentage of synonymful synsets than *BN*, with a gain of 18 points. The best results are obtained by *MT&BN* with an average gain of 6 points with respect to *MT*. Although the *BN* dictionary has limited impact on overall synsets coverage (with a gain of 2.4 points, as shown in *Experiment 2*), *BN* improves synonym words coverage by an average of 6 points, which can have a significant impact on mapping selection with majority voting. For instance, when compared to *MT*, the *MT&BN* dictionary improves the percentage of correct mappings in *TopOne* selection for the synonymful synsets that are fully covered by 4.6 points, as shown in Figure 5.5(b). Covering more synonym words belonging to a synonymful synset, not only improves synsets coverage, but also makes the mapping selection step easier. Thus, integrating more resources for translation can be advantageous in the mapping selection tasks as well.

Observation 12. For synonymful synsets, the larger the number of synonym words covered by translations, the easier the mapping selection task is.

5.4.3 Experiment 3: Coverage and Correctness of Translations vs Concept Specialization

I recall that a synset is not covered when none of the words of its equivalent synset in the target language is returned by its translation. In other words, when a synset is not covered, the correct match cannot be found among the set of candidate matches found by translation. This analysis further helps in the exploration of the problem of synset coverage by investigating 1) the impact of domain specificity on synset coverage, and 2) the possibility of improving the coverage by expanding the set of found candidate matches with synsets similar to the ones retrieved by translations.

To investigate if non covered synsets can be characterized to some extent based on their specificity, I use two different methods to characterize specificity: the domain labels associated with synsets in *WordNet Domains* [11], e.g., biology, animals, and so on; the position that synsets occupy in the semantic hierarchies, e.g., synsets that occur as leaf nodes in the hypernym hierarchies.

I consider a synset which is associated with a domain label in Wordnet domains as domain-specific, i.e., every label different from "*Factoum*" (i.e., general, or non specified domain). For every wordnet, the percentage of domain specific synsets that are not covered by MT dictionary is shown in Figure 5.6. For example, I found that, on average, only 36% of the non covered synsets with the MT dictionary are labeled as "*Factoum*". The rest of the non covered synsets (64%) are distributed over different domains (with biology, animals, person, plants, and geography as the most frequent ones). These findings consolidate the ones discussed in *Experiment 1*: monosemous words, which do often express specific concepts, were found to be less covered than polysemous words, which often express more general concepts.

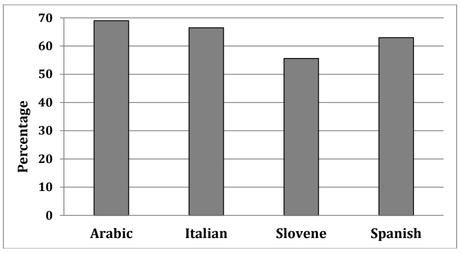


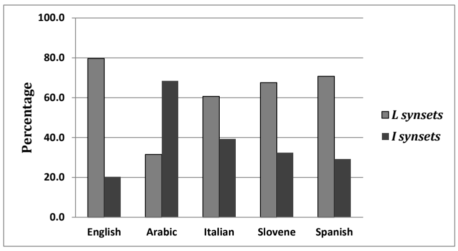
FIGURE 5.6: Percentage of *domain specific* synsets that are not covered by MT

Observation 13. Domain-specific concepts have less coverage, by machine translation tools, than general concepts.

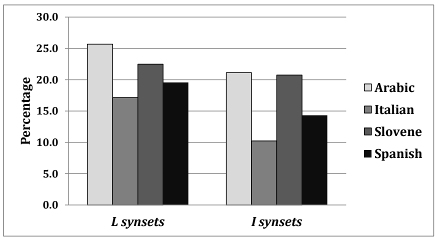
With the same intent, I consider how synsets not covered by translations are distributed in the semantic hierarchy defined by the hypernym/hyponym relation. In this context, I consider leaf synsets (called *Lsynsets*) as the most specific synsets, while intermediate synsets (called *Isynsets*), i.e., synsets occurring in other positions in the hierarchy, are considered to be more generic. I consider only a subset of synsets, i.e., *nominal synsets*, whose hierarchical structure is well-established in the English wordnet. In particular, to determine the position of a source synset I consider the position of its equivalent synset in the English WordNet, by using the mappings existing between the wordnets.

Figure 5.7(a) reports the percentage of *Lsynsets* and *Isynsets* for every wordnet. I can notice that most of the wordnets have more leaf synsets than intermediate synsets, with the exception of the Arabic wordnet. This exception can be explained by the strategy used for the construction of this wordnet and by its relatively small size. The construction of the Arabic wordnet [97], which is based on the expand model paradigm introduced in the EuroWordNet project [115], was initiated by the translation of the *core concepts* of the English WordNet [19], and was, thereafter, extended to other concepts. The core concepts (over 5 thousands) are often assumed to be common across different cultures and languages, and are often intermediate synsets.

Figure 5.7(b) reports the percentage of *Lsynsets* and *Isynsets* that are not covered with MT dictionary for each wordnet. The average percentage of nominal *Lsynsets* and *Isynsets* not covered with the MT dictionary is 21.1% and 16.6%, respectively. Table 5.12 reports, for every wordnet, the distribution of nominal *Lsynsets* vs *Isynsets*, grouped by synset category. I can notice that *Lsynsets*



(a)



(b)

FIGURE 5.7: Percentage of Leaf synsets ($Lsynsets$) and Intermediate synsets ($Isynsets$) (a) in the gold standards (b) in the non-covered synsets

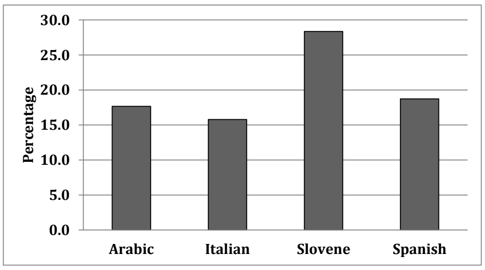


FIGURE 5.8: Neighbour synset coverage for non-covered synset

are more likely to be not covered than $Isynsets$, and that a large number of non-covered synsets consists of synonymless synsets.

Moreover, I would like to evaluate if, for non covered synsets, translations return candidate matches that are at least *semantically similar* to their equivalent synsets in the target language. Neighbor synsets (i.e., hypernyms, hyponyms, and siblings) are usually considered similar in many wordnet and graph based similarity measures [85]. Inspired by work presented in [96], one could consider establishing a weighted mapping between a synset in a source language and synsets in the target language, such that the weight represents the degree of similarity between the source and the target synset. In this experiment I define similar synsets as one being either hyponym/hypernym or siblings of the other one. As shown in Figure 5.8, the average percentage of synsets not covered with MT for which at least one synset similar to the equivalent synset is found among the candidate matches is 20.1%. This is consistent with the intuition that machine translation systems provide translations which (implicitly) capture a more coarse-grained sense specification than the fine-grained sense specification encoded in the wordnets. In fact,

TABLE 5.12: Distribution of leaf and intermediate (non-)covered synsets by category

Synsets	Arabic				Italian			
	Non-covered		Covered		Non-Covered		Covered	
	Leaf	Inter	Leaf	Inter	Leaf	Inter	Leaf	Inter
M-MWS	9.1	16.0	35.0	39.9	5.3	1.5	71.6	21.5
M-OWS	14.3	27.1	27.0	31.6	11.4	4.0	60.3	24.3
MIX	3.7	8.5	19.6	68.2	4.1	2.1	44.6	49.3
P-MWS	2.7	5.2	17.9	74.2	4.0	2.6	34.4	59.0
P-OWS	8.8	13.1	19.7	58.4	11.3	5.5	44.1	39.0
ALL	7.8	14.4	23.5	54.3	8.6	3.6	52.1	35.7

Synsets	Slovene				Spanish			
	Non-covered		Covered		Non-Covered		Covered	
	Leaf	Inter	Leaf	Inter	Leaf	Inter	Leaf	Inter
M-MWS	10.9	4.7	60.0	24.3	8.0	1.1	82.5	8.4
M-OWS	13.2	4.6	70.4	11.9	17.8	4.2	64.9	13.1
MIX	11.3	7.1	38.1	43.5	9.6	2.4	48.7	39.3
P-MWS	13.0	8.3	28.1	50.6	7.1	3.3	28.3	61.3
P-OWS	22.3	10.2	35.3	32.2	14.1	7.0	37.0	41.8
ALL	15.2	6.7	52.4	25.7	13.8	4.2	56.9	25.0

it was observed that WordNet is sometimes too fine-grained even for human judges to agree [56].

Observation 14. For a significant percentage of non covered synsets (20.1%, on average), machine translation tools return synsets that are at least similar to their equivalent synsets in the target language.

Based on this observation, the candidate match retrieval step can be modified so as to include among the candidate matches also synsets similar to the ones retrieved by translation. This approach has been followed by several cross-lingual ontology matching systems [44, 39, 28]. However, expanding the set of considered candidate matches has the disadvantage of increasing the difficulty of mapping selection task. The results of this analyses suggest that the expansion of the candidate matches set is a technique that could be applied only to particular categories of source synsets, e.g., to *synonymless leaf* synsets. This could provide a system (or a user, in an interactive matching settings) with a greater ability to map synsets that are less likely to be covered by translations, without increasing the number of candidate matches for every source synset, e.g., for synsets that have *distinguishable monosemous* candidate matches (see *Observation* 8).

5.4.4 Lessons Learned

In this section I summarize the main results and findings of this study and highlight some potential future directions.

A general conclusion that I draw from this study is that machine translation tools and multilingual knowledge bases return useful translations for a very large number of concepts. Thus, translations provide a valuable support for candidate match retrieval in cross-lingual ontology matching, covering from a minimum of 75.2% to a maximum of 89.9% synsets in the four languages other than English considered in this study. If I consider that BabelNet also incorporates translations derived from mappings in the Open Multilingual Wordnet [14] (which have been excluded in this study because they have been used as gold standards), this coverage is expected to even increase for several resource-rich languages covered by this wordnet. In addition, these experiments suggest that translations can be helpful, to a more limited extent, and for selected categories of synsets, also in the mapping selection task.

Concisely, the main results of these experiments suggest that:

- monosemous concepts (i.e., concepts that have only monosemous words) are considered to be more domain-specific;
- combining translation resources improves the quality of results;
- automatic translations perform poorer on domain-specific concepts than on domain-independent ones;
- synonymful synsets have higher coverage than synomymless synsets;
- most, but not all, monosemous concepts can be mapped confidently even with simple selection methods (e.g., translation-based majority voting);
- mappings involving polysemous but synonymless synsets are harder to filter out within the mapping selection task;
- the more the coverage for synonym words (in synonymful synsets), the easier is the mapping selection task.

Compared to previous systems, which used machine translation tools considering only one translation direction, in this study I built dictionaries that cover both translation directions by including reverse translations. This technique has been shown to significantly improve the coverage of translations. In practice, candidate matches can be found for a larger number of input concepts, thus increasing the upper-bound recall for cross-lingual ontology matching systems. As a promising future research direction, one may try to further improve coverage by considering additional information available in machine translation tools like Google Translate (e.g., reverse translation synonym-like sets, part-of-speech tagged translations, and translation scores). Such additional information can increase not only the upper-bound recall, but also the precision, if adequately used in the matching selection step. For example, one may compare the words returned by reverse translations with an input source synset, e.g., by using the translation-correctness measure (Eq.5.3). The translation with higher translation-correctness could be given a higher weight in the selection step.

The selection of a correct mapping from a set of candidate matches still remains a difficult task, in particular when *contextual knowledge* cannot be used to disambiguate the meaning of the concepts. However, the findings of the study suggest several research directions that can mitigate this problem.

On one hand, the simple baseline selection method based on majority voting used in these experiments should be overcome by more sophisticated methods. For example, in a recent work, I define a lexical similarity measure based on evidence collected from translations and I run a local similarity optimization algorithm to improve the assignments between source and target concepts [3]. In future work, I would like to leverage the analysis of mapping selection difficulty as dependent on the lexical characterisation of source and target concepts (e.g., polysemous vs. monosemous concepts, or synonymless vs. synonymful synsets) discussed in this paper. I plan to investigate matching algorithms that could adapt their behavior based on the category of the source synset and its candidate matches.

On the other hand, some cross-lingual mappings may still be hard to decide upon using a fully automatic approach. Thus, I would like to investigate in the cross-lingual ontology matching domain, the adoption of semi-automatic matching methods. A web application could be used to solve difficult cross-lingual matching tasks, as the one proposed to match short service descriptions in different languages [83]. Beyond this, interactive matching processes that aggregate inputs given by a multiplicity of users, either experts [26] or crowd workers [100] seem particularly promising in large cross-lingual matching tasks. The findings of this paper are particularly useful for similar approaches because they can help to decide on which mappings the user inputs are more valuable (e.g., polysemous and synonymless concepts). Overall I plan to follow the latter research directions to use a map model to ease the construction of a "lexical-semantic" ontology in the context of the Arabic Ontology Project [60, 6], which also motivated the study presented in this paper.

5.5 Conclusions

In this study I have investigated the effectiveness of automatic translations derived from a state-of-the-art machine translation tool (Google Translate) and a state-of-the-art multilingual ontology (BabelNet) to support cross-lingual ontology mapping. To perform this analysis I used four very large repositories of cross-lingual mappings, which include mappings from wordnets in four different languages to the English WordNet. Effectiveness of automatic translations is analyzed in terms of coverage and correctness. One key contribution of this study, besides the scale of the experiments, is the analysis of the effectiveness of automatic translations for specific categories of synsets.

For example, I found that automatic translations achieve lower coverage for domain specific concepts. As another example, I found that the amount of monosemous words that are correctly translated into polysemous words in another language is not negligible: cross-lingual ontology mapping methods that use the monosemous word heuristic may lead to include a several wrong mappings in an alignment. At a coarse grain, the analyses suggest that automatic translations are capable of covering a large number of word senses, in particular when more resources (e.g., Google Translate and BebelNet) and translation strategies (i.e., the reverse translations of Google Translate) are integrated. On the other hand, automatic translations are correct only to a limited extent, at least when compared to translations derived from manually mapped wordnets.

The analyses discussed in this chapter inspired the definition of a cross-lingual similarity measure for lexical ontologies 6. A natural subsequent step is to further utilize the study outcomes in cross-lingual mapping systems. One promising research direction is to define adaptive mapping methods such that different strategies are used depending on the lexical characterization of the source concepts. For example, one could integrate interactive mapping methods or crowdsourcing approaches to decide about a subset of the mappings, which are estimated to be particularly difficult to map. Another research direction that I plan to investigate is a method to estimate of concept ambiguity in small ontologies that do not explicitly contain synonyms, e.g., by matching them against wordnets. Such a method would help us to use adaptive cross-lingual mapping methods on axiomatic ontologies or other lexically-poor data sources, e.g., web tables.

Chapter 6

Cross-Lingual Lexical Mapping Method

6.1 Chapter Overview

As introduced in Chapter 1, data on the Web is lexicalized in different languages world-wide. Different lexical resources such as ontologies, thesauri, and dictionaries have been shown to provide valuable support for data integration in language-specific applications [33, 15, 105, 95, 28]. Merging these resources is a crucial task to support the integration of information lexicalized in different languages [46]. In this chapter, I present a cross-lingual lexical mapping method to map lexically-rich resources, i.e., resources that associate each concept with a set of synonym words.

In Section 6.2, I overview the proposed mapping method. In Section 6.3, I discuss a purely lexical, *translation-based similarity measure* (TSM), which uses translations as evidence to compute the similarity between the source synsets and the candidate matches. In Section 6.4, I explain a disambiguation technique based on a *Local Similarity Optimization Algorithm* (LSOA), which leverages the similarity between the source synsets and the candidate matches in order to decide the best mappings, and discuss how to obtain the final alignment. Further, an experimental design to evaluate the performance of the proposed method, baseline methods, evaluation measures, and obtained results are discussed in Section 6.5. In Section 6.6, I discuss the impact of the structural evidence and how smoothly it can be incorporated along with the proposed method. Finally, in Section 6.7, I draw main conclusions and highlight potential future directions.

The scientific accomplishment presented in this chapter; Sections 6.2, 6.3, 6.4, and 6.5, have been published in [3].

6.2 Approach Overview

Given a source synset s in one language, this method tries to find the best match for s among the set of target synsets in another language. The principles of this method, which considers only the concepts' lexicalization, can be sketched as follows. I collect a large set of candidate matches for a source synset by translating each of its synonym words (without trying to disambiguate their meaning) using a machine translation system and a large multilingual knowledge resource. The words returned by the translation are used as evidence in order to rank the candidate matches by computing their similarity with the source synset, using a novel similarity measure (explained in detail in Section 6.3). I use a disambiguation technique to select and decide the best mappings, i.e., to assign the best matches to each source synset using a novel *Local Similarity Optimization Algorithm* (LSOA), discussed in Section 6.4.

To evaluate the proposed method, I use cross-lingual mappings manually established by lexicographers between four wordnets (Arabic, Italian, Slovene and Spanish) and the English WordNet (described in Chapter 2, in Section 2.5.1). Translations are obtained using external multilingual resources. In the experiments I use two resources as sources of translation, which are frequently adopted in the research field: *Google Translate* and *BabelNet*, as explained in Section 2.4.1.

Following the cross-lingual mapping task definition in Section 2.5, I consider collections of synsets (e.g., a dictionary or thesaurus) rather than structured ontologies. Mappings can represent different relations between source and target synsets. I consider only equivalence mappings, i.e., mappings which specify that a source synset and a target synset have equivalent meaning. Mappings can be defined as triples $\langle s, t, w \rangle$, where s and t are synsets respectively of the source and target collections, and $w \in [0; 1]$ is a *weight* representing the confidence that the synsets s and t have an equivalent meaning. The set of mappings returned by the cross-lingual mapping task is also called *alignment*. Following the assumption in most of the stat-of-the-art work [103, 111], I also assume that the cardinality of an alignment containing equivalence mappings is 1:1, i.e., each source synset is mapped to at most one target synset.

Cross-lingual mapping tasks incorporate three main components (explained in Section 2.5). These components are used in an algorithm that selects the mappings to include in the final alignment; the algorithm includes a disambiguation technique based on a *Local Similarity Optimization Algorithm*. I explain in detail the algorithm in Section 6.4. Intuitively, I want to include in the alignment a mapping between every source synset and their best candidate matches. However, it should be noticed that it is not always possible to include a mapping for every synset (a problem common also to baseline methods). As discussed in Section 5.4.2, different candidate matches may be evaluated to be equally good for a source synset based on the available evidence, i.e., a *tie* occurs among a set of top-ranked matches; in this case, the mapping for s is **undecidable**, and no mapping for s is included in the final alignment. I call this **TopOne** selection approach, which preserves the 1:1 cardinality of the alignment by electing at most one mapping for each mapping. However, sometimes, it might be useful to consider a brave selection approach, such that every mapping that is considered among the best matches is included in the alignment. I call the latter approach **TopSet** and I will use it for evaluation purposes in Section 6.5.

6.3 Translation-based Similarity Measure

In this section, I introduce (to the best of my knowledge) a novel *cross-lingual similarity measure* (for short, TSM), which is inspired by the classification-based interpretation of mappings' semantics [9]. I use the F_1-measure to compute the similarity between a source synset s and a target synset t. The F_1-measure is the harmonic mean of recall and precision measures.

Precision is defined as the number of words returned by translation of s (Eq. 2.1) that occur in t, over the total number of words returned by translation of s.

$$Precision : P(s,t) = \frac{|\{sTrans_D^{L_2}(s)\} \cap t|}{|\{sTrans_D^{L_2}(s)\}|}. \tag{6.1}$$

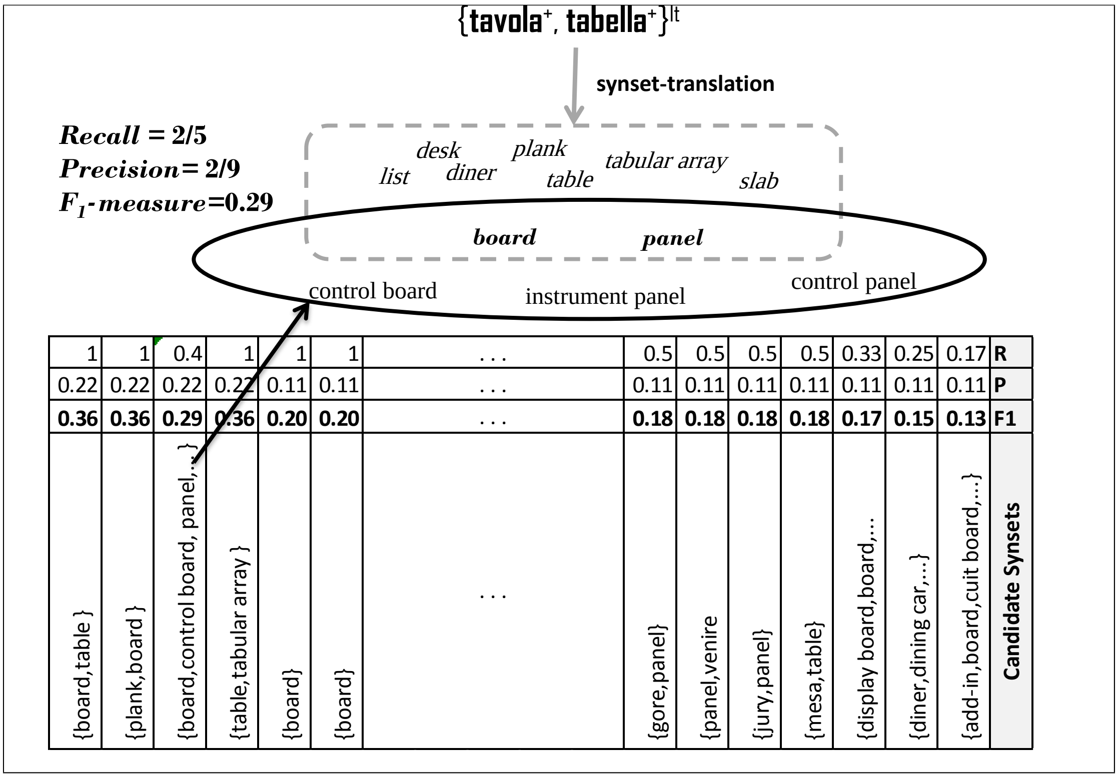


FIGURE 6.1: Classification-based similarity measure

Recall is defined as the number of words returned by translation of s that occur in t, over the total number of words in t.

$$Recall : R(s,t) = \frac{|\{sTrans_D^{L_2}(s)\} \cap t|}{|t|}. \tag{6.2}$$

The weight of a mapping between s and t synsets is given as F_1-measure:

$$F_1{-}measure : F_1(s,t) = 2 \times \frac{P(s,t) \times R(s,t))}{(P(s,t) + R(s,t))}. \tag{6.3}$$

For a more in-depth analysis of the semantics of cross-lingual mappings, details have been provided in Chapter 4.

For example, Figure 6.1 shows the TSM computed between the Italian synset $\{$*tavola*$^+$, *tabella*$\}$ and its candidate matches in English. For instance, the weight for the target synset $\{$*board*$^+$, *control board,...*$\}$ is computed as follows: recall $R = \frac{2}{5}$, precision $P = \frac{2}{9}$, and $F_1 = 0.29$.

6.4 Mapping by Merging Locally Optimal Assignments

Given a ranked set of candidate matches, a simple approach for mapping selection is to map a source synset to its best match (when a tie does not occur). However, this solution does not consider that the decision about a mapping can be influenced by other mappings that are relevant for the decision, e.g., mappings between other source synsets and the same candidate matches. In previous work on mono-lingual ontology [27], the mapping selection problem has been viewed as an instance of the Assignment Problem [18]: mappings included in the alignment are those that maximize the similarity of the whole alignment. Finding a *global optimal solution* for an assignment problem requires combinatorial algorithms (e.g., the Hungarian Method [65]), which are costly in terms of memory usage and execution times and cannot be applied to match thousands, or dozens of thousands of synsets (e.g., wordnets). Even the efficient solution proposed in [27] can be hardly applied to resources of the size considered in this thesis (Section 2.5.1), because it runs on a bipartite graph representing the whole alignment.

The mapping selection method proposed in this chapter is based on merging locally optimal assignments computed for each source synset. For each source synset s, a locally optimal assignment between a (small) set S of source synsets lexically related to s, and a (small) set T of target synsets, is computed; these *pre-alignment* mappings are stored with their weights. Every pre-alignment mapping $\langle s, t, w \rangle$ such that s is not pre-aligned to any other mapping is included in the final alignment. If a source synset s is mapped to more than one target synsets that have the same weight, the match for s is undecided. In the following Section 6.4.1, I present the *Local Similarity Optimization Algorithm* (LSOA) where I explain in detail how I compute each locally optimal assignment.

6.4.1 Local Similarity Optimization Algorithm

In this section, I explain the Local Similarity Optimization Algorithm (or, LSOA).

Initially, for each source synset a set of candidate matches is retrieved, in the candidate match retrieval task (discussed in Section 2.5). Then, the locally optimal assignments is computed using the Hungarian method, which operate over a *weight*

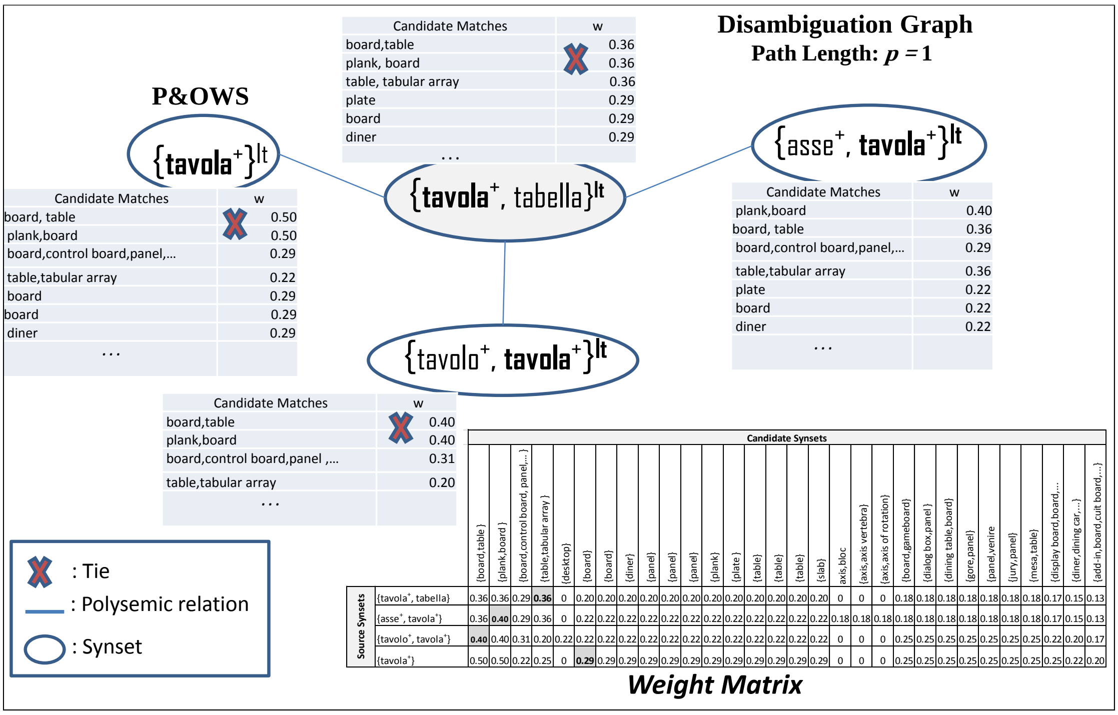


FIGURE 6.2: Illustrative example of the LSOA

matrix. A weight matrix $M = S \times T$ is created for each source synset s such that: S contains s and a set of synsets lexically related to s, and T consists of all the candidate matches for synsets in S. The matrix is created by iteratively expanding S and T.

The set S is populated using a *disambiguation context* for the synset s, i.e., source synsets for which it is likely to find mappings in conflict with s. To determine the disambiguation context of a synset s, a graph $G = (V, E)$, called **disambiguation graph**, is constructed. Each node $v \in V$ is a synset, which contains a set of words $\{w_1, w_2, ..., w_n\}$. Two synsets are connected by an edge iff they have at least one word in common, which is called *polysemic relation*. The relation is represented by edges in the disambiguation graph, because it is based on polysemic words shared by synsets. The graph is built iteratively from s, by performing a breadth-first search over the polysemic relation with a depth limit. The depth is defined by a maximum path length p, which is given as input to the algorithm. Figure 6.2 shows an excerpt of the disambiguation graph for the Italian synset $\{tavola^+, tabella\}$ with $p = 1$.

Once the disambiguation graph is built, and the candidate matches for each synset in the graph are retrieved, the weight matrix is populated. The value of a cell represents the weight assigned to the synsets s_i and t_i, which is computed using

TSM (as described in Section 6.3) and is defined as follows: $cell[i;j] = w_{ij}$ if the weight (similarity) assigned to the synsets s_i and t_i, $cell[i;j] = 0$ if $\nexists w_{ij}$. After populating the weight matrix, a locally optimal assignment can be computed using the Hungarian algorithm. The execution of the algorithm consists of a series of iterative steps that generate mappings with a maximum weight.

Observe that the similarity between a source synset s and every target synset t in the disambiguation graph, which is not a candidate match for s, is equal to 0. In fact, whenever the similarity between a source synset s and a target synset t is different from 0, t will appear among the candidate matches for s by definition. For this reason LSOA always finds mappings among the candidate matches of source synsets.

Figure 6.2 shows an illustrative example of a disambiguation graph built from the Italian synset $\{$*tavola*$^+$, *tabella*$\}$; each node is attached with a descending order candidate matches based on their similarity (STM) weights. Three nodes have undecidable mappings because of ties. With LSOA the selection task considers the synsets in the disambiguation graph. Once LSOA assigns the target synsets to the synset $\{$*asse*$^+$, *tavola*$^+\}$. This reduces the likelihood that the English synset $\{$*plank*$^+$, *board*$^+\}$ will be the target synset of $\{$*tavola*$^+$, *tabella*$\}$ or $\{$*tavola*$^+$, *tabella*$\}$. Then, the English synsets $\{$*board*$^+$, *table*$^+\}$ and $\{$*table*$^+$, *tabella*$\}$, are assigned to the source synsets $\{$*tavolo*$^+$, *tavola*$^+\}$, and $\{$*tavola*$^+$, *tabella*$\}$, respectively, in order to maximize the total similarity. Based on the *Observation 9*, discussed in Section 5.4.2, mappings involving polysemous but synonymless synsets (for short, $P\&OWS$) are harder to filter out within the mapping selection task; for the synset $\{$*tavola*$^+\}$ a tie occurs, nevertheless, the algorithm will randomly select one target synset out of the $TopSet$ candidate matches. $P\&OWS$ synsets can be correctly selected only if all the conflicting synsets (synsets in the graph) are correctly disambiguated, and the target synset is not a $P\&OWS$ synset.

6.5 Experiments

The main goal of the experiments performed is to evaluate the performance of the proposed matching system in mapping a lexical resource in a language other than English to the English WordNet, i.e., in the Cross-Lingual Mapping (CLM) tasks. To this end, I will compare mappings found with the proposed matcher to mappings available in gold standard alignments. In addition, I want to evaluate the completeness of the lexicalizations that can be found in the target language by using alignments automatically created with the proposed method. In order to evaluate the extent to which the proposed mapping method can be helpful to enrich a multilingual semantic resource, I will use gold standard alignments. These alignments help to evaluate the richness of the lexicon found using the proposed mapping method when compared to the lexicon found using different approaches proposed in related work [86].

I evaluate the matching method considering three different configurations: using the similarity measure TSM with selection based on *TopOne* mappings, i.e., ranking the similarity weights without applying the optimization method, TSM with selection based on LSOA at p=1, TSM with selection based on LSOA at p=2. I compare these approaches with two baseline approaches: Monosemous Words Heuristic, and Majority Voting. I run the experiment over a system with $8GB$ RAM and $1.33GHz$ core. I used MT dictionary (Section 2.4.1) for translation in the above configurations.

Next, I describe the experimental design and the evaluation measures. Further, I describe the baseline approaches. Finally, I discuss the experiments results.

6.5.1 Evaluation Measures

Three experiments have been conducted, where in the first two experiments I focus on the cross-lingual mapping task, and I evaluate the accuracy of the mappings found with different matchers against gold standards. I use the well-known performance measures Precision, Recall, and F_1-Measure, to quantify the performance of the matchers. Results are computed as macro-average measures [114]. The recall, precision, and F_1-measure are expressed in terms of percentage, ranging from 0% - 100%. In the third experiment, I focus on the task of enriching multilingual semantic resources with lexicalizations in one language. These lexicalizations are obtained with a matcher that align a large lexical resource in one language, e.g., English WordNet, to a lexically-rich resource in another language (e.g., an Italian dictionary). To this end, I compute the recall of the lexicalizations found with a given method against the lexicalizations defined in the gold standards. This measure shows that for a source synset, how many words in the target synset, in the target language, are returned with the enrichment method. Observe that, since I cannot expect that such lexicalizations are complete for every language, measuring precision is not meaningful.

6.5.2 Baseline Approaches

I compared the performance of the matching approach with two baseline approaches as described below.

6.5.2.1 Monosemous Words Heuristic

The monosemous words heuristic (MWH) selection method is based on the assumption that a monosemous word in a source language is translated into a monosemous word in a target language. MWH applies only on monosemous-synsets. Given a monosemous-synset s, the monosemous words are translated, then word translation is used in sense-lookup. If the same candidate synset is returned, i.e, they are all monosemous word in the target language, it will be selected as the correspondent synset.

MWH has been adopted in several state-of-the-art systems, e.g., [86], as heuristic because it has the advantage that several correct mappings can be selected in a simple way. However, synsets in different languages, which have an equivalent meaning, can be lexicalized in different types of words (Section 5.2). For example, the Italian monosemous-synset {forchetta} is mapped to the English synset {fork$^+$}, which is a $P\&OWS$ synset. In Figure 2.2, the Italian monosemous word *"tabella"* is translated into three words in English; two are polysemous words. Therefore, the correct target synsets cannot be selected with MWH. If I consider the translation of its synonym word, the word *"tavola$^+$"*, I can collect additional evidence that can be used to select the correspondent synset.

6.5.2.2 Majority Voting

The majority voting (MV) selection method considers the *multiset union* of the candidate matches that are returned by every sense-lookup task for a source synset. Given a source synset s, the set of candidate matches can be defined as the set $T_s = \{t_i^n | t_i \in sense(w) \wedge w \in sTrans_D^{L_2}(s)\}$, where n indicates the number that

TABLE 6.1: Upper-bound gold standards

Translation	Synsets	Arabic	Italian	Slovene	Spanish
BN_{core}	Monosemous	44.2	70.6	55.2	38.0
	All	37.4	62.5	44.2	44.7
BN	Monosemous	46.0	72.0	58.3	39.1
	All	51.3	72.8	52.0	49.0
MT	Monosemous	74.1	88.3	78.3	79.3
	All	77.7	87.6	72.4	79.7

the target synset t_i is returned by sense-lookup task. MV determines the best target synset by selecting the most frequent synset.

For example, in Figure 2.2 the target synsets that have the highest frequency are $\{board^+, table^+\}^3$ and $\{table^+, tabular\ array\}^3$; each appear three times in the candidate matches set. These are *undecidable* mappings; i.e., a *tie* has occurred. Nevertheless, two synsets are nominated as best target synsets that include the correct match, instead of considering all candidate matches in T_s.

6.5.3 Experiment 1: Quality of the Alignment

In this experiment I, evaluate the performance of the alignments found with every matcher. When a matcher is not able to decide about a mapping due to a tie (undecidable mappings), the mapping is considered incorrect. I compare the three configurations TSM, TSM+LSOA$_{p=1}$, and TSM+LSOA$_{p=2}$ with the baseline approaches MWH, MV.

I first evaluate the performance measures against the whole gold standard. Then I perform the same evaluation using a subset of the gold standard, which is defined by the mappings such that the target synset is among the candidate matches found for the source synset by the candidate match retrieval phase. I call this subset **upper-bound** gold standard. I define the upper-bound of a matcher as the *synset-coverage*; which evaluates if the translation function of the source synset returns at least one word of the correspondent synset, as defined in [86, 5](Section 2.4). This is motivated by the observation that every selection method described in this chapter can select only matches that are found in the candidate match retrieval step. Thus candidate match retrieval defines an upper-bound for the recall of every method. By evaluating the alignment found by the selection methods with this subset of the gold standard, I want to evaluate the performance of the selection methods in isolation from the limits imposed by the candidate match retrieval phase. In Table 6.1, I report the upper-bounds of the matchers; as a

Table 6.2: Exp 1: Matcher performance w.r.t gold standards and upper-bounds

		MWH			MV			TSM		
Lang.	Synsets	R	P	F1	R	P	F1	R	P	F1
					w.r.t. gold standards					
Ar	Monos.	14.0	44.1	21.3	25.4	53.6	34.4	29.2	47.5	36.2
	All	–	–	–	17.7	37.2	24.0	20.5	32.7	25.2
It	Monos.	45.8	**95.4**	60.3	47.5	82.0	60.2	55.2	75.0	63.6
	All	–	–	–	32.1	63.8	42.7	40.0	58.4	47.5
Slv	Monos.	48.9	89.6	63.3	58.2	91.6	71.2	61.4	83.1	70.6
	All	–	–	–	34.5	73.7	47.0	38.9	62.1	47.9
Es	Monos.	38.1	82.6	52.2	49.4	83.8	62.2	60.3	79.3	68.5
	All	–	–	–	31.9	60.7	41.9	37.9	53.0	44.2
					w.r.t. upper-bounds					
Ar	Monos.	16.4	44.1	23.9	34.3	53.6	41.8	39.4	47.5	43.1
	All	–	–	–	22.7	37.2	28.2	26.3	32.7	29.2
It	Monos.	46.9	**95.4**	62.9	53.8	82.0	65.0	62.5	75.0	68.2
	All	–	–	–	36.7	63.8	46.6	45.7	58.4	51.3
Slv	Monos.	55.9	89.6	68.8	74.4	91.6	82.1	78.4	83.1	80.7
	All	–	–	–	47.6	73.7	57.8	53.8	62.1	57.7
Es	Monos.	46.8	82.6	59.8	62.4	83.8	71.5	68.5	79.2	73.5
	All	–	–	–	40.1	60.7	48.3	44.1	57.6	49.9

		TSM+LSOA$_{p=1}$			TSM+LSOA$_{p=2}$		
Lang.	Synsets	R	P	F1	R	P	F1
				w.r.t. gold standards			
Ar	Monos.	45.2	73.0	55.8	**52.1**	**73.6**	**61.0**
	All	29.9	**69.6**	41.8	**38.7**	66.8	**49.0**
It	Monos.	66.4	89.6	76.3	**73.3**	90.0	**80.8**
	All	43.7	**88.3**	58.5	**55.4**	85.3	**67.2**
Slv	Monos.	65.1	**91.9**	76.2	**69.0**	90.5	**78.3**
	All	36.8	**91.1**	52.4	**45.1**	85.6	**59.0**
Es	Monos.	60.8	**89.0**	72.3	**64.6**	88.7	**74.7**
	All	37.8	**87.3**	52.7	**44.2**	83.6	**57.8**
				w.r.t. upper-bounds			
Ar	Monos.	61.0	73.0	66.5	**70.3**	**73.6**	**71.9**
	All	38.4	**69.6**	49.5	**49.7**	66.8	**57.0**
It	Monos.	75.2	89.6	81.8	**83.0**	90.0	**86.4**
	All	50.0	**88.3**	63.8	**63.3**	85.3	**72.7**
Slv	Monos.	83.2	**91.9**	87.3	**88.2**	90.5	**89.4**
	All	50.8	**91.1**	65.2	**62.3**	85.6	**72.1**
Es	Monos.	76.7	**89.0**	82.4	**81.4**	88.7	**84.9**
	All	47.4	**87.3**	61.4	**55.5**	83.6	**66.7**

relative number of covered synsets in the gold standards with different translation resources. I report the synset-coverage for the monosemous-synsets and all synsets in the gold standards. Table 6.2 reports the performance measures compared to mappings in the all gold standard datasets and upper-bounds. For LSOA I report results obtained with disambiguation graph with path length p, for values $p = 1$ and $p = 2$; experimentally at $p = 3$ the system fails to perform all mapping tasks; on average execution time is 15 and 115 minuets at $p = 1$ and $p = 2$, respectively.

6.5.4 Experiment 2: Recall of Undecidable Mappings

The aim of this experiment is to understand if correct mappings can be found in the set of *TopSet* candidate matches. In the selection methods used so far source synsets, for which a tie occurs, are not mapped to any target synset, i.e., their mapping is undecided. However, if the correct mapping can be found among these top-ranked candidate matches, and if this set is reasonably small, one could use interactive matching approaches, e.g., based on crowdsourcing [26], to decide about the mappings. For this reason, in this experiment, I consider a correctly retrieved mapping if it appears in a set of top-ranked candidate matches, with cardinality less or equal than $\mathcal{C}$, for different values of $\mathcal{C}$. Also in this case, I compare the selection methods both with the entire gold standards and upper-bounds datasets. Table 6.3 reports the recall for MV, TSM and $\text{LSOA}_{p=1}$ methods with *TopSet* sets at cardinality $\leq \mathcal{C}$; for $\mathcal{C}= 1$ (*TopOne*), 5, and *all* the *TopSet* set. In *TopSet* setting, a mapping is considered correct for a source synset, whenever the correct match for the synset is included in the set of its top-ranked candidate matches. Observe that every mapping that is counted as correct in *TopOne* setting, is also counted as correct in *TopSet* setting. I report results for LSOA only at $p = 1$ as it has reasonable execution time. Mappings with MWH are not selected based on a similarity weight, so it is not included.

6.5.5 Experiment 3: Multilingual Lexical Enrichment

In this experiment, I compare the lexicalizations of languages other than English found by mapping non-English lexical resources to English WordNet, with three matching configurations: MV, TSM, and $\text{LSOA}_{p=1}$ and compared to three sets of synsets, which I refer to as **baseline-synsets**. As baseline-synsets, I use BabelNet synsets [86]: BN_{core} and BN (described in Section 2.4.1); and a third set is synsets constructed with MT where correspondent synsets are obtained directly with the translation of the source synsets, i.e., with out employing any selection method. I compare the lexicalization of synsets returned with the different matchers at *TopOne* setting, with the lexicalization in the baseline-synsets. The intuition behind the baseline-synsets, on one hand, is to highlight the advantage of using selection methods w.r.t direct translation (i.e., MT), which presents a naive simple matching approach; whereas, on the other hand, is to compare the returned synsets

TABLE 6.3: Exp 2: Recall of *TopSet* selection at different cardinality

	w.r.t. gold standards									
		MV			TSM			TSM+LSOA$_{p=1}$		
	Synset	≤ 1	≤ 5	*all*	≤ 1	≤ 5	*all*	≤ 1	≤ 5	*all*
Ar	Monos.	25.4	40.2	48.8	29.2	38.2	40.5	45.2	58.2	62.2
	All	17.7	29.9	39.2	20.5	29.2	32.8	29.9	47.6	53.9
It	Monos.	47.5	69.7	76.8	55.2	67.0	68.5	66.4	78.1	80.5
	All	32.1	54.6	64.9	40.0	53.5	56.1	43.7	64.5	69.7
Slv	Monos.	58.2	69.5	74.6	61.4	66.5	66.9	65.1	73.1	75.7
	All	34.5	50.3	62.1	38.9	49.3	51.7	36.8	55.1	63.9
Es	Monos.	49.4	62.5	66.4	54.6	62.0	62.8	60.8	67.3	70.1
	All	31.9	45.2	52.8	36.7	46.1	48.2	37.8	52.0	61.5
	w.r.t. upper-bounds									
		MV			TSM			TSM+LSOA$_{p=1}$		
	Synset	≤ 1	≤ 5	*all*	≤ 1	≤ 5	*all*	≤ 1	≤ 5	*all*
Ar	Monos.	34.3	54.3	65.8	39.4	51.6	54.7	61.0	78.6	83.9
	All	22.7	38.3	50.4	26.3	37.5	42.2	38.4	61.1	69.2
It	Monos.	53.8	78.9	87.0	62.5	75.8	77.5	75.2	88.4	91.1
	All	36.7	62.4	74.1	45.7	61.2	64.1	50.0	73.7	79.6
Slv	Monos.	74.4	88.8	95.3	78.4	84.9	85.5	83.2	93.4	96.7
	All	47.6	69.5	85.8	53.8	68.2	71.5	50.8	76.2	88.3
Es	Monos.	62.4	78.8	83.7	68.9	78.2	79.2	76.7	84.9	88.4
	All	40.1	56.8	66.3	46.0	57.9	60.5	47.4	65.3	77.2

TABLE 6.4: Exp 3: Recall of the enrichment task

En to	MT	BN$_{core}$	BN	MV	TSM	TSM+LSOA$_{p=1}$
Ar	59.7	21.5	29.7	56.3	**78.6**	65.8
It	68.2	45.3	53.5	66.8	**80.4**	70.5
Slv	53.8	43.0	47.3	58.7	**70.3**	60.3
Es	65.6	30.5	39.9	58.7	**78.5**	71.2

w.r.t synsets in BabelNet, which is arguably considered as the largest state-of-the-art multilingual knowledge system.

6.5.6 Discussion

Results in Table 6.2 show that the performance of MV, TSM, and LSOA for the monosemous-synsets outperform the MHW approach. MWH has higher precision than the other methods, since large number of the monosemous-synsets are synonymless synsets [5]. However, MWH has the lower recall which can be explained by two main observations. First, MWH single-out the monosemous words from its synonym words, which excludes several potential candidate matches. Second, monosemous words that are considered as specific domain words, are covered by the translation resource less than the polysemous words, which are considered as frequent and generic domain words to a great extent [5].

Table 6.2 shows that MV and TSM results are comparable; MV achieves higher precision measures than TSM, whereas TSM achieves higher recall measures than MV. An interesting direction, is to study how to join the force of both methods, e.g., linear weighted combination matcher [27]. Observe that in Table 6.2 the precision measures are the same for gold standards and upper-bounds datasets.

LSOA method reports the best measures w.r.t the other methods. LSOA take advantage over MV and TSM by considering mappings in the disambiguation graph. In Figure 6.2 the Italian synsets $\{tavola^{+},\ tabella\}$ could not be disambiguated with MV or TSM due to a tie; which resulted in undecidable mappings. With LSOA the selection task of the synset $\{board^{+},\ tavola^{+}\}$ increases the likelihood that the correct mapping of the synsets $\{asse^{+},\ tavola^{+}\}$ and $\{tavola^{+},\ tabella\}$ is selected. With LSOA I get closer to the upper-bounds for the monosemous-synsets (Table 6.1), again since mapping tasks are not performed in isolation of other mapping tasks, and synonymful synsets that contains some monosemous and polysemous words are better disambiguated with the context graph by resolving the undecidable mappings.

In Table 6.2 recall measures still have considerable margins w.r.t upper-bounds in Table 6.1. In fact, Table 6.2 reports the performance with the $TopOne$ selection setting, that is, I did not consider cases when ties occur; whereas Table 6.3 details this case. Observe that recall measures are remarkably enhanced when I consider the ties (undecided mappings).

Table 6.4 shows that the synsets that are obtained with MV, TSM, and LSOA methods are significantly lexically-richer than baseline-synsets. Observe that, TSM has the highest recall, which is due to the fact that it selects the target synsets in a such way that it maximizes the recall, whereas LSOA tries to obtain a trade-off between precision and recall.

6.6 Combine Lexical and Structural Evidences

So far, I have presented a pure lexical mapping method for mapping concepts in different languages, however, structural information for source and target ontologies might be available, even partially, in different mapping scenarios. In this section, I want to evaluate the quality of the obtained alignment w.r.t gold standards when structural and lexical evidence are used together in order to decide the best mappings. By doing this, I show the significance of incorporating structural evidence with the proposed method. In other words, I show the usefulness of using structural and lexical evidence in achieving more correct mappings.

To do so, I use the Descendant's Similarity Inheritance (DSI) method [29]. Next, in Section 6.6.1, I overview DSI method. Further, in Section 6.6.2, I present an experimental design to evaluate the quality of the obtained alignment when using DSI, and discuss the main results.

6.6.1 Structural-based Method

The Descendant's Similarity Inheritance (DSI) method [29] is an automatic alignment method that uses the structure of the ontology for contextual information. DSI reconfigures the *base similarity* between the concepts based on the similarity of their parent concepts. Base similarity, is the very first step to establish initial mappings between the concepts of the source ontology and the concepts of the target ontology. These initial mappings will be a starting point for the DSI methods. Next, I present the details of the DSI method as described in [29].

Given a source ontolgoy S and a target ontology T, the DSI method reconfigures similarity between a concept C in S and a concept C' in T, defined as $DSI\_sim(C, C')$. In order to determine $DSI\_sim(C, C')$ the following steps are performed sequentially:

- Let $path\_len\_root(C)$ be the number of edges between the concept C in S and the root of the ontology S.
- Let $parent_i(C)$ be the i^{th} concept from the concept C to the root of the source ontology S, where $0 \leq i \leq path\_len\_root(C)$. Similarly define $parent_i(C')$ with respect to T.
- Define MCP as the *main contribution percentage*, which is the fraction of the similarity measure between C and C' that will be used in determining the overall $DSI\_sim(C, C')$.
- Compute $DSI\_sim(C, C')$ as follows:

$$MCP * base\_sim(C, C') + \frac{2(1-MCP)}{n(n+1)} \sum_{i=1}^{n} (n+1-i) base\_sim(parent_i(C), parent_i(C'))) \tag{6.4}$$

where n = $min(path\_len\_root(C), path\_len\_root(C'))$.

The main characteristic of the DSI method is that it allows for the parent and in general for any ancestor of a concept to play a role in the identification of the concept. Intuitively, the parent of a concept should contribute more to the identity of the concept than its grandparent. This is achieved by setting a relatively high value to main contribution percentage (MCP) factor. The grandparent concept contributes more than the great grandparent, and so on, until the root is reached. I set the value of the MCP factor to 75% as suggested in [29].

6.6.2 Experiment: Impact of Structural Evidence

In this experiment, first I evaluate the performance of the obtained alignments w.r.t gold standards and upper-bounds. Then, I evaluate the improvement in recall of undecidable mappings. For DSI method, as a base similarity measure, I use the translation-based similarity measure (TSM), explained in Section 6.3. I compare results obtained in Experiment 1 (Section 6.5.3) and Experiment 2 (Section 6.5.4), with three configurations: using DSI with selection based on *TopOne* mappings, i.e., ranking the similarity weights without applying the optimization method, using DSI with selection based on LSOA at p=1 (DSI+LSOA$_{p=1}$, i.e., first run DSI, then sequentially perform LSOA), using DSI with selection based on LSOA at p=2 (DSI+LSOA$_{p=2}$).

The experimental results are reported in Table 6.5 and Table 6.6: Table 6.5 reports the performance measures compared to mappings in the all gold standard datasets and upper-bounds. Table 6.6 reports the recall for TSM, DSI and DSI+LSOA$_{p=1}$ methods with *TopSet* sets at cardinality $\leq \mathcal{C}$; for $\mathcal{C}$= 1 (*TopOne*), 5, and *all* the *TopSet* set.

By utilizing the structural evidence by using the DSI method, which is used to re-configure (refine) the similarity between concepts, the experimental results demonstrate its usefulness in determining more accurate similarity measures between concepts, hence, more correct mappings. This happens since several ties (undecidable mappings) have been resolved after reconfiguring the similarity weights TSM

TABLE 6.5: Matcher performance w.r.t gold standards and upper-bounds: using structural&lexical deviance

Lang.	Synsets	MWH R	MWH P	MWH F1	MV R	MV P	MV F1	TSM R	TSM P	TSM F1	DSI R	DSI P	DSI F1
w.r.t. gold standards													
Arabic	Monos.	14.0	44.1	21.3	25.4	53.6	34.4	29.2	47.5	36.2	48.3	54.5	51.2
	All	–	–	–	17.7	37.2	24.0	20.5	32.7	25.2	37.8	43.2	40.3
Italian	Monos.	44.1	95.4	60.3	47.5	82.0	60.2	55.2	75.0	63.6	68.5	74.9	71.6
	All	–	–	–	32.1	63.8	42.7	40.0	58.4	47.5	54.0	60.6	57.1
Slovene	Monos.	48.9	89.6	63.3	58.2	91.6	71.2	61.4	83.1	70.6	77.0	82.9	79.8
	All	–	–	–	34.5	73.7	47.0	38.9	62.1	47.9	55.3	64.9	59.7
Spanish	Monos.	38.1	82.6	52.2	49.4	83.8	62.2	60.3	79.3	68.5	68.7	78.7	73.4
	All	–	–	–	31.9	60.7	41.9	37.9	53.0	44.2	48.6	57.3	52.6
w.r.t. upper-bounds													
Arabic	Monos.	16.4	44.1	23.9	34.3	53.6	41.8	39.4	47.5	43.1	55.8	54.5	55.2
	All	–	–	–	22.7	37.2	28.2	26.3	32.7	29.2	44.3	43.2	43.7
Italian	Monos.	46.9	95.4	62.9	53.8	82.0	65.0	62.5	75.0	68.2	73.0	74.9	74.0
	All	–	–	–	36.7	63.8	46.6	45.7	58.4	51.3	59.2	60.6	59.9
Slovene	Monos.	55.9	89.6	68.8	74.4	91.6	82.1	78.4	83.1	80.7	86.1	82.9	84.4
	All	–	–	–	47.6	73.7	57.8	53.8	62.1	57.7	69.3	64.9	67.0
Spanish	Monos.	46.8	82.6	59.8	62.4	83.8	71.5	68.9	79.3	73.8	75.8	78.7	77.2
	All	–	–	–	40.1	60.7	48.3	46.0	55.0	50.1	56.4	57.3	56.8

Lang.	Synsets	TSM+LSOA$_{p=1}$ R	TSM+LSOA$_{p=1}$ P	TSM+LSOA$_{p=1}$ F1	TSM+LSOA$_{p=2}$ R	TSM+LSOA$_{p=2}$ P	TSM+LSOA$_{p=2}$ F1	DSI+LSOA$_{p=1}$ R	DSI+LSOA$_{p=1}$ P	DSI+LSOA$_{p=1}$ F1	DSI+LSOA$_{p=2}$ R	DSI+LSOA$_{p=2}$ P	DSI+LSOA$_{p=2}$ F1
w.r.t. gold standards													
Arabic	Monos.	45.2	73.0	55.8	52.1	73.6	61.0	52.6	74.4	61.6	55.6	72.2	62.8
	All	29.9	69.6	41.8	38.7	66.8	49.0	39.7	71.7	51.1	42.4	64.7	51.2
Italian	Monos.	66.4	89.6	76.3	73.3	90.0	80.8	70.2	89.5	78.7	74.6	89.3	81.3
	All	43.7	88.3	58.5	55.4	85.3	67.2	51.1	88.7	64.8	57.5	84.9	68.6
Slovene	Monos.	65.1	91.9	76.2	69.0	90.5	78.3	68.9	91.8	78.7	70.2	90.1	78.9
	All	36.8	91.1	52.4	45.1	85.6	59.0	44.5	91.7	59.9	48.5	85.0	61.7
Spanish	Monos.	60.8	89.0	72.3	64.6	88.7	74.7	63.2	88.9	73.9	66.0	88.1	75.4
	All	37.8	87.3	52.7	44.2	83.6	57.8	44.1	87.7	58.7	47.7	83.1	60.6
w.r.t. upper-bounds													
Arabic	Monos.	61.0	73.0	66.5	70.3	73.6	71.9	68.4	74.4	71.3	71.4	72.2	71.8
	All	38.4	69.6	49.5	49.7	66.8	57.0	48.2	71.7	57.6	50.9	64.7	57.0
Italian	Monos.	75.2	89.6	81.8	83.0	90.0	86.4	79.0	89.5	83.9	83.3	89.3	86.2
	All	50.0	88.3	63.8	63.3	85.3	72.7	57.3	88.7	69.6	63.7	84.9	72.8
Slovene	Monos.	83.2	91.9	87.3	88.2	90.5	89.4	87.0	91.8	89.3	88.3	90.1	89.2
	All	50.8	91.1	65.2	62.3	85.6	72.1	58.5	91.7	71.5	62.5	85.0	72.0
Spanish	Monos.	76.7	89.0	82.4	81.4	88.7	84.9	79.1	88.9	83.7	81.9	88.1	84.9
	All	47.4	87.3	61.4	55.5	83.6	66.7	53.8	87.7	66.7	57.3	83.1	67.8

between the source and target concepts. The quality of alignment when using DSI outperforms the other configurations, which do not incorporate the structural information, as reported in Table 6.5 and 6.6. This in fact, confirms the hypotheses that incorporating more contextual knowledge (the structural evidence in this case) improves the quality of alignment.

TABLE 6.6: Recall of *TopSet* selection at different cardinality: using structural&lexical deviance

	Synset	TSM ≤ 1	TSM ≤ 5	TSM *all*	DSI ≤ 1	DSI ≤ 5	DSI *all*	STM+LSOA$_{p=1}$ ≤ 1	STM+LSOA$_{p=1}$ ≤ 5	STM+LSOA$_{p=1}$ *all*	DSI+LSOA$_{p=1}$ ≤ 1	DSI+LSOA$_{p=1}$ ≤ 5	DSI+LSOA$_{p=1}$ *all*
w.r.t. gold standards													
Ar	Monos.	29.2	38.2	40.5	48.3	52.6	52.9	45.2	58.2	62.2	52.6	60.1	64.1
	All	20.5	29.2	32.8	37.8	41.8	42.4	29.9	47.6	53.9	39.7	49.9	56.2
It	Monos.	55.2	67.0	68.5	68.5	73.2	73.5	66.4	78.1	80.5	70.2	78.9	81.3
	All	40.0	53.5	56.1	54.0	59.0	59.5	43.7	64.5	69.7	51.1	65.9	71.0
Slv	Monos.	61.4	66.5	66.9	77.0	78.9	78.9	65.1	73.1	75.7	68.9	73.8	76.4
	All	38.9	49.3	51.7	55.3	59.5	60.4	36.8	55.1	63.9	44.5	56.2	64.9
Es	Monos.	54.6	62.0	62.8	68.7	72.0	72.2	60.8	67.3	70.1	63.2	68.0	70.7
	All	36.7	46.1	48.2	48.6	53.2	53.8	37.8	52.0	61.5	44.1	53.5	63.0
w.r.t. upper-bounds													
Ar	Monos.	39.4	51.6	54.7	48.3	52.6	52.9	61.0	78.6	83.9	68.4	80.5	85.8
	All	26.3	37.5	42.2	37.8	41.8	42.4	38.4	61.1	69.2	48.2	63.4	71.5
It	Monos.	62.5	75.8	77.5	68.5	73.2	73.5	75.2	88.4	91.1	79.0	89.3	91.9
	All	45.7	61.2	64.1	54.0	59.0	59.5	50.0	73.7	79.6	57.3	75.1	81.0
Slv	Monos.	78.4	84.9	85.5	77.0	78.9	78.9	83.2	93.4	96.7	87.0	94.2	97.4
	All	53.8	68.2	71.5	55.3	59.5	60.4	50.8	76.2	88.3	58.5	77.2	89.3
Es	Monos.	68.9	78.2	79.2	75.8	79.4	79.6	76.7	84.9	88.4	79.1	85.6	89.0
	All	46.0	57.9	60.5	56.4	61.7	62.4	47.4	65.3	77.2	53.8	66.8	78.7

6.7 Conclusions

In this chapter, I presented a novel cross-lingual lexical selection approach. I introduced a new cross-lingual similarity measure, which is inspired by a classification-based mapping semantics to measure the similarity between two synsets lexicalized in different languages. I used a disambiguation technique to assign the best match to each source synset using a novel local similarity optimization algorithm. I evaluated the approach using wordnets in four different languages, which have been manually mapped to the English wordnet. Experiments showed that the approach significantly improves the quality of an automatically generated alignment even when applied after purely lexical, recall-oriented and efficient techniques for candidate match retrieval and similarity evaluation. Moreover, experiments showed that by using a merge model first, and enriching synsets' lexicalizations using the found mappings with the proposed approach afterwards, one can obtain richer lexicalizations for the source synsets. Moreover, the experimental results showed that, using structural evidence and lexical evidence from translation together has improved the quality of alignments.

For future work, I plan to enhance the recall of the proposed approach, as well as extend the Hungarian algorithm so that it does not stop after the generation of the (first) best mapping but continues to the generation of the best $top - k$ mappings, similar to work presented in [12]. Further, comparing the TSM with other similarity methods should be also investigated, e.g, vector based similarity measures [35]. Not only to compare their performance, but also investigate how to join the force of different methods, e.g., linear weighted combination matcher [27]. Another interesting direction is to evaluate the proposed method in more systematic matching scenarios, like OAEI. This poses several challenges like richness of concepts (discussed in Section 3.4). One way is to enrich the concepts (e.g., as described in Section 6.5.5), and then decide the best mappings using the disambiguation algorithm LSOA.

Chapter 7

Interactive Mapping Application

7.1 Chapter Overview

As described in the previous chapters, the manual cross-lingual mapping requires considerable effort, which makes it unfeasible at a large scale [31]. Automatic cross-lingual mapping methods can be used either to compute mappings automatically (e.g., [3]), even at the price of accuracy, or to support semi-automatic mapping workflows by recommending mappings to lexicographers.

With automatic matching system, different candidate matches may be evaluated to be equally good for a source synset based on the available evidence, i.e., a *tie* occurs among a set of top-ranked matches; in this case, the mapping for this source synset is undecidable, and no mapping for this synset is included in the final alignment. However, if the correct mapping can be found among top-ranked candidate matches, and if this set is reasonably small, one could use interactive mapping approaches, e.g., inspired by crowdsourcing model [26], to decide about the mappings. In Section 6.5.4, the experimental results, in which I investigated such scenarios, showed an improvement in recall. For instance, more than 15% increase in the recall can be achieved if the top five ranked matches are considered.

Based on this finding, in this Section, I present ICLM (Interactive Cross-lingual Mapping) application, which is a semi-automatic mapping approach that supports feedback provided by multiple lexicographers. In This approach; first an alignment is computed using automatic matching methods and then allows for the lexicographers (called henceforth *users*) to validate them.

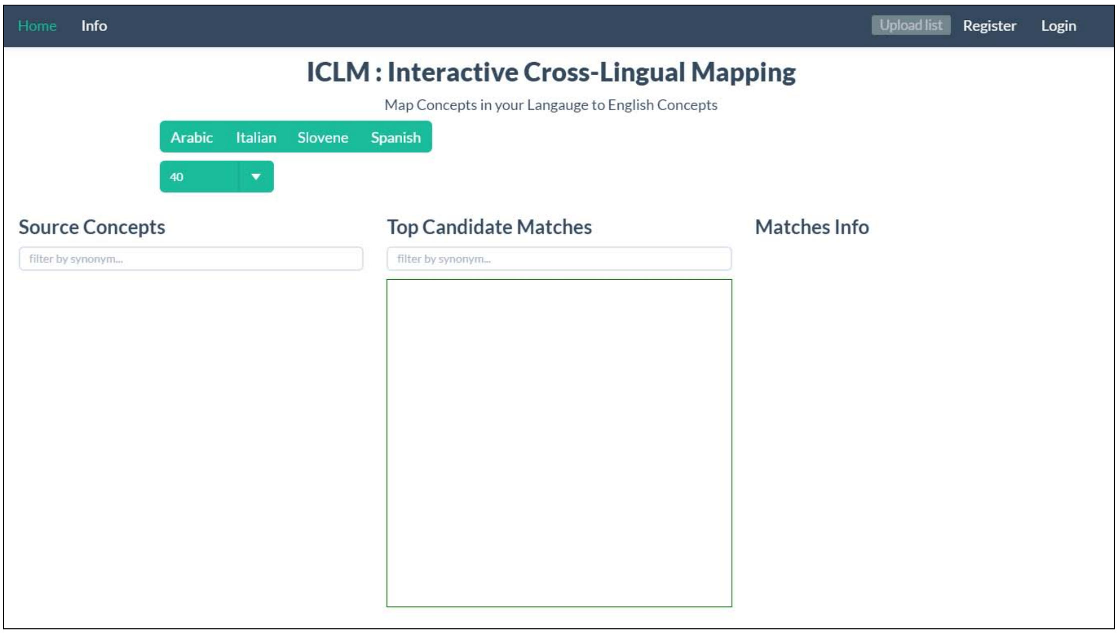


FIGURE 7.1: ICLM home page

The rest of this chapter is organized as follows: in Section 7.2, I overview ICLM and provide more insights on the functionalities provided by the approach and its the Web GUI. Further, in Section 7.3, I describe the key elements of the proposed approach: strategies to estimate the validation efforts required from users for each source concept. In Section 7.4, I discuss the conducted experiment: the dataset, a model for the evaluation of the quality of the mappings and users effort, and the results. Finally, in Section 7.6, I draw some conclusions and describe future work. In what follows, I consider the scenario of mapping Arabic concepts to concepts in the English WordNet.

7.2 ICLM Overview

ICLM[1], Interactive Cross-Lingual Mapping, is a Web application that supports a semi-automatic mapping procedure aiming at speeding up and improving an automatically generated alignment. ICLM tries to reduce the users efforts in validating cross-lingual concept mappings. ICLM distributes the mapping (validation) tasks on users based on the mapping tasks' difficulties, i.e., in the validation process, ICLM defines the number of users based on the difficulty of the mapping selection task. ICLM estimates the difficulties of the mapping selection tasks based

[1] http://193.204.59.21:1982/iclm/

on lexical characteristics (as discussed in Chapter 5) of concepts under evaluation and on how confident the automatic matching algorithm is, i.e., ICLM estimates the task difficulty and accordingly estimates the expected users effort (number of users to validate a mapping task). Figure 7.1 shows the home page of ICLM.

Initially the source concepts will be automatically matched against the target concepts using automatic matching methods, TSM+LSOA (described in Section 6.4). Then, the system estimates the mapping selection tasks difficulty; and accordingly defines the number of users to validate each task (explained in details in Section 7.3). In this way, ICLM distributes the mapping tasks over some users based on the estimated efforts, unlike pure crowdsourcing models, e.g., [100], which equally assign the same number of users for every task. The user is free to select any source concept from the source list. Once the user identifies the potentially correct candidate match, he choses one relationship that reflects his decision (described in details in Section 7.4).

Since more than one user is involved, ICLM uses a *consensus-based* approach to decide whether a mapping belongs to the final alignment. Similar to previous work [26], ICLM uses a consensus model based on simple majority vote, where V is an odd number of validations considered sufficient to decide by majority (ICLM does not require that all the users validate each mapping task); thus, *minimum consensus*, $\mu = \lfloor (V/2) + 1 \rfloor$, is the minimum number of similar vote that is needed to make a final decision on a mapping. For example, if $V = 5$ is the number of validations considered sufficient to decide by majority, a final decision on a mapping can be taken when $\mu = 3$ similar vote are assigned to a mapping by the users.

Every mapping that obtains the minimum consensus of votes will be confirmed, i.e., included in the final alignment, and will be removed from the source concepts list for this specific user. Once the user finishes his task, a confirmation message is sent, and the corresponding task is removed from the source list. However, other users may still find it, for instance, if the minimum consensus of votes has not reached. After each validation task, ICLM updates the source list until the whole mappings are validated. In this way, ICLM reduces and saves more of users efforts. Cases where agreement (the minimum consensus of votes) is not achieved, the match which has the highest rank and received more votes will be included in the final alignment. Otherwise, it will not be included in the final alignment.

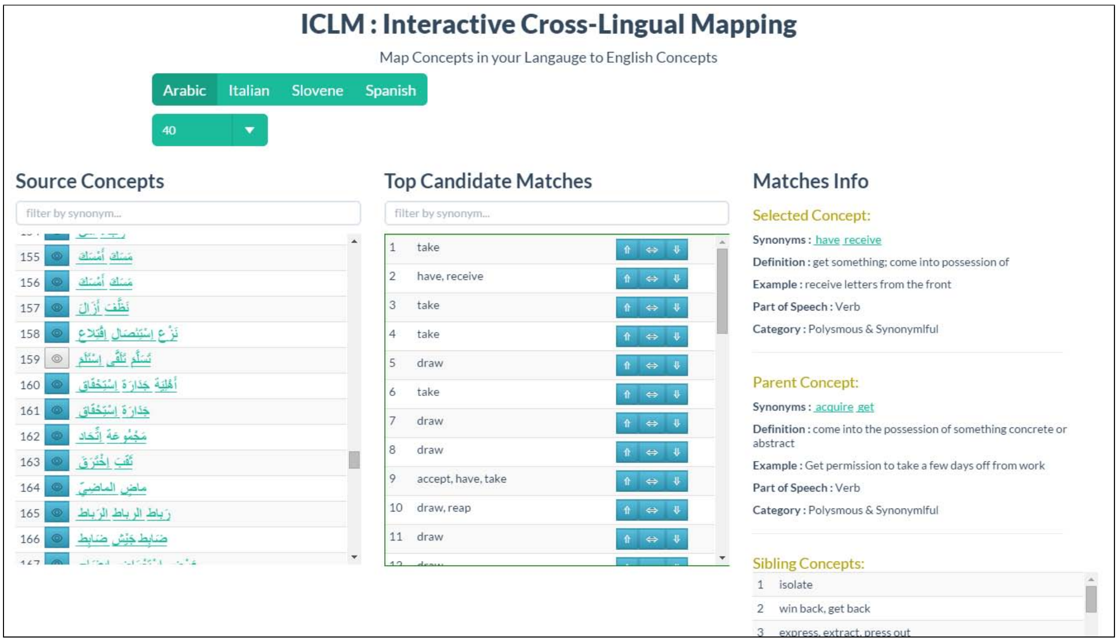


FIGURE 7.2: ICLM: supports user with useful details

Observe that the agreement factor (i.e., the minimum consensus of votes) can be tuned in a favor to increase the mapping accuracy by increasing this factor. However, this comes at a price of increasing the users effort. Furthermore, different agreement strategies can be adopted. For example, mapping tasks will be confirmed only if a given number of users have agreed without controlling the number of users who are validating the mapping tasks. This of course will increase the users efforts. One may consider feedback reconciliation models more sophisticated than majority or weighted majority voting, for example, tournament solutions [26]. This would be an interesting direction as a future work to explore.

Next, I provide more insight on the functionalities provided by the application and on the Web GUI[2]. Figure 7.2 illustrates ICLM's functionalities. Before the users start use the application, the source concepts are automatically matched to the English concept using TSM+LSOA (as described in Section 6.4). The first step that the ICLM user should perform is to *Register* and *Login*, so to enable all the validation functionalities. After that, he select the respective language of concepts to be mapped to concepts in the English WordNet. The user is now able to explore the source concepts by scrolling the whole list of concepts (*Source Concepts*) or by performing a keyword-based search (see Figure 7.2). Next, the user selects a source concepts and ICLM retrieves a list of candidate English concepts (*Top Candidate*

[2]ICLM Web GUI has been adapted from CroSeR Web GUI [83].

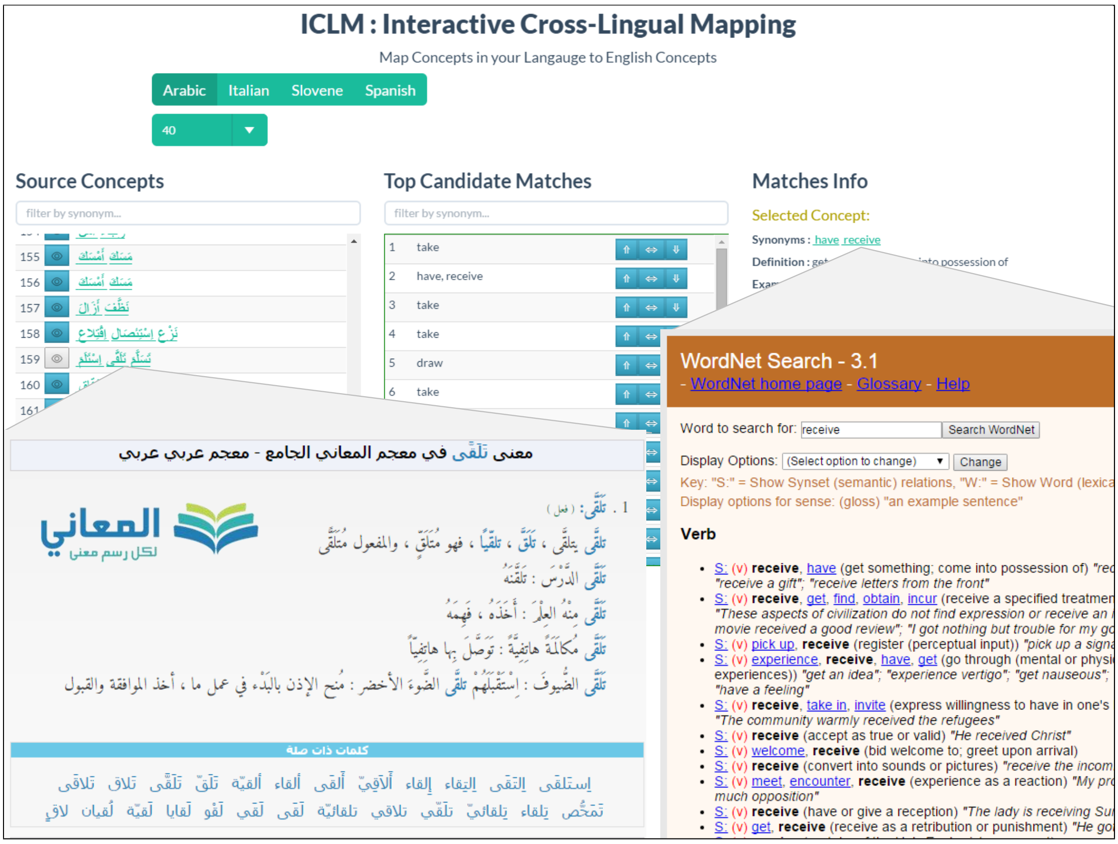


FIGURE 7.3: ICLM's details snapshot

Matches), that are potentially equivalent. The number of retrieved matches is configurable by the user through the GUI (e.g., the 40 top-ranked matches).

Since, the connection between the source concept and the target concept could be not straightforward by simply comparing concepts' lexicalization, a user can then select a candidate match and look at further details (*Matches Info*) directly gathered from the English WordNet. Moreover, the user can click the source and target concepts lexicalization to get further information, as depicted in Figure 7.3. For instance, the user will be able to access an online glossary for the source language[3], as well as navigate through the semantic hierarchy of the English WordNet via the online English WordNet website[4]. Finally, the user can switch on the feedback mode of ICLM which would store the selected relation between the source concept and the English concept.

[3] In the current implementation, for Arabic ICLM uses Al-maany glossary: http://www.almaany.com/ar/dict/ar-ar/, which returns all possible senses of a given word (i.e, the word is not disambiguated).

[4] http://wordnetweb.princeton.edu/perl/webwn

For each mapping (validation) task ICLM logs the users' activities: the elapsed time of each mapping task; and users' navigation activities (accessing the external resources: the glossary or the online English WordNet). In this way, I can evaluate the effectiveness and usability of ICLM, discussed in Section 7.4.

7.3 Estimating the Required Number of Validations

The basic idea behind ICLM is to reduce the users' effort in validating a pre-defined alignment, and thus speeding up the mapping process. In order to estimate the mapping selection tasks' difficulties, so as to estimate the required efforts (number of users required for validation), ICLM leverages the lexical characteristics of concepts under evaluation, where the confidence of the candidate matches is based on the lexical based disambiguation algorithm LSOA (Section 6.4). ICLM considers the following features of concepts under evaluation to estimate the validation tasks difficulties:

- Ambiguity of lexicalization:
  - Monosemous words: words that have only one sense (meaning).
  - Polysemous words: words that have two or more senses.
    Polysemous words are more difficult to disambiguate than the monosemous one when contextual knowledge is limited.
- Synonym-richness:
  - Synonymless : a concept that is lexicalized with one word.
  - Synonymful: a concept that is lexicalized with many words.
    The more the coverage for synonym words (in synonymful synsets), the easier is the mapping selection task.
- Uncertainty in the selection Step: matches which can be obtained by an automatic cross-lingual mapping systems, e.g., LSOA (Section 6.3), in which the candidate matches are ranked based on their similarity degree with the source concepts.

- TopOne: if there exists a unique top-ranked candidate match for the source synset.
- TopSet: if there exists a set of top-ranked matches, i.e., a unique top-voted candidate match does not exist for the source synset.

Three **validation strategies** have been defined in ICLM: Low, Mid, and High levels of difficulty. In each level ($l := \{L, M, H\}$), different number of users are required to validate the mapping of each source concept, i.e., the number of validation tasks that are considered sufficient to decide by majority ($V^l \geq 0$). The validation strategies levels are as follow:

- **Low-difficulty**: V^L validation tasks are required.
- **Mid-difficulty**: V^M validation tasks are required.
- **High-difficulty**: V^H validation tasks are required.

Each level can have a different agreement factor, i.e., the minimum consensus of votes. Accordingly, different configurations can be considered as trade-offs between mappings accuracy and users efforts. For instances, $V^l = 0$ suggests that mappings will be directly included in the final alignment without any feedback (validation). An increase in the value of V^l means increasing the users efforts and the mapping accuracy, under the assumption that users are expected to identify the correct relation with out introducing errors (which is not always the case). Observe that, more validation strategy levels can be introduced, based on application requirements.

ICLM applies the following rules in order to select the respective validation strategy, i.e., define the number of validation tasks that are considered sufficient to decide by majority:

- **Low-difficulty**: if a monosemous synset is under evaluation and TopOne candidate match exist, *OR* if a synonymfull synset is under evaluation and TopOne candidate match exist.
- **Mid-difficulty**: if a source synset does not have a TopOne match.
- **High-difficulty**: if a source synset is polysemous and synonymless (P&OWS).

7.4 Experiment

The goal of this experiment is to investigate the effectiveness of ICLM in suggesting good candidate matches; not only for equivalent relation but also for relationships different from the equivalent relation, i.e., specific/general concepts. This experiment also investigates the quality of the classification approach, which is used to define the validation strategies, hence, estimate the number of validation tasks (i.e., number of users). In other words, the experiment investigates if the estimated difficulties of the mapping selection tasks confirms the observations concluded from the study in Chapter 5.

I evaluate the performance of ICLM considering two different configurations; based on the number of validation tasks assigned to each validation difficulty level: *BRAVE strategy* (V^L=0, V^M=1, V^H=3), the Low-difficulty tasks will be included into the final alignment without validation; and *CAUTIOUS strategy* (V^L=1, V^M=3, V^H=5), every task will be validated by some users.

I evaluate the performance of the alignments found with every configuration against a gold standard. I use the well-know performance measures of Precision, Recall, and F_1-measure, to quantify the performance of the alignments. I compare the two configurations with an alignment automatically obtained using the configuration TSM+LSOA (discussed in Section 6.4), i.e, without validation (V^L=0, V^M=0, V^H=0).

Next, I describe the experimental settings. the gold standard and the users involved in the validation tasks. Further, I describe the validation tasks (steps that users follows). Finally, I report the main results of the experiment.

In this experiment six bilingual speakers, from different background: geography, computer since, law, medicine, management, and engineering are asked to link a set of Arabic concepts, taken from the Arabic wordnet (ArWN) [97], to concepts in the English WordNet by using ICLM. Users are undergraduate students (2), postgraduate students (2), and doctorates (2). These users are knowledgeable about ICLM and its goals. For each source concept ICLM retrieves the set of candidate matches, which are ranked based on their similarity with the source concept (discussed in Section 6.4).

Dataset and sampling criteria: I randomly selected 250 concepts from the Arabic wordnet, such that certain condition are satisfied. The concepts are selected to

Table 7.1: Sample dataset: distribution by category

Category	M&OWS	M&MWS	MIX	P&OWS	P&MWS	**Total**
ArWN (%)	19.3	13.4	24.7	21.2	21.4	100.0
Sample	48	36	62	52	52	250
Decidable	24	18	31	26	26	125
Undecidable	24	18	31	26	26	125

Table 7.2: Sample dataset: distribution by validation difficulty

Validation strategy	Low-difficulty	Mid-difficulty	High-difficulty	Total
Sample (number)	99	99	52	250
Sample (%)	39.6	39.6	20.8	100.0

Table 7.3: Sample dataset: distribution by synonym words

Synonym words	1	2	3	4	5	6	7	9	Total
Sample(number)	101	72	47	18	6	2	3	1	250

Table 7.4: Sample dataset: distribution by word type

Word Type	noun	verb	adjective	adverb	Total
ArWN(%)	68.8	24.3	5.9	1.0	100.0
Sample (number)	166	62	19	3	250
Sample(%)	66.4	24.8	7.6	1.2	100.0

Table 7.5: Sample dataset: distribution by TopSet cardinality

cardinality of TopSet	[4-10]	[11-20]	[21-40]	Total
Sample(number)	93	24	8	125

Table 7.6: Sample dataset: distribution by (noun) concepts specialization

Position in the hierarchy	[1-3]	[4-6]	[7-9]	[10-12]	[13-15]	Total
Sample(number)	2	47	88	26	2	166

reflect a uniform distribution (w.r.t the gold standard, see first row in Table 7.1)) of concepts category as well as tasks difficulty; the following factors are considered while selecting the concepts: decidable vs undecidable mappings, the number of synonym words in a source concept, the type (part of speech) of concepts lexicalization, the size of the top-ranked matches in the undecidable mappings, and the position of concepts in the semantic hierarchy (concepts' specialization). Tables 7.1, 7.2, 7.3, 7.4, 7.6, and 7.5 report these details.

The validation tasks are processed as follows: After registration, a user can access and start validating the matches. The following instructions (guidelines) are provided to the users:

- *register* to the system and *login*;

- select the respective language (*Arabic*) of the source concept list;
- select the *Full List* of candidate matches;
- select one of the source concepts from the Arabic concept (*Source Concepts*);
- evaluate the list of candidate matches (*Top Candidate Matches*);
- if the lexicalization (synonyms) of a candidate matches is not sufficient to validate the mapping, click on the candidate match for getting more details. In the *Matches Info* side one can find more useful details, which includes definitions, examples, and neighbor (parent and sibling) concepts. These information are navigable to the online English WordNet. Similarly, an online Arabic glossary (Al-maany glossary website) is also accessible and linked to each source synonym word (see Figure 7.3). Use the full-text search if a correct candidate match does not appear in the top positions;
- once identified the potentially correct candidate match, choose one of the following relationships:
  - General ($\Uparrow$): the candidate concept is more generic with respect to the source concept;
  - Equivalent ($\Leftrightarrow$): the candidate concept is equivalent to the source concept;
  - Specific ($\Downarrow$): the candidate concept is more specific with respect to the source concept;
- select another concept from the source list until all the concept have been evaluated.

7.5 Results and Discussion

Table 7.7 reports the performance measures for the three configurations. Precision (P) measures how many selected relation are correct w.r.t the gold standard. Recall (R) measures how many correct relations are selected w.r.t the gold standard. F_1-measure is the harmonic mean of the two measures. The first row reports the performance of selecting the equivalent relations, while the second row reports if also specific or general relations are also correctly selected.

TABLE 7.7: Performance results: different validation configurations

Configuration	TSM+LSOA $V^L = V^M = V^H = 0$			BRAVE validation $V^L = 0, V^M = 1, V^H = 3$			CAUTIOUS validation $V^L = 1, V^M = 3, V^H = 5$		
Relation	R	P	F1	R	P	F1	R	P	F1
Equivalent	50.0	50.0	50.0	59.6	66.7	63.0	68.4	71.8	70.1
Equivalent, Specific, General	-	-	-	67.2	67.6	67.4	79.6	73.4	76.4
# Required validation	-			[203-255]			[453-650]		
# Preformed validation	-			233			556		
# Avg. time/validation (sec)	-			97			89		

The third row reports the required number of validations: the lower bound refers to the minimum number of validations, which happen if a consensus agreement occurs for each source concept; whereas the upper bound refers to the maximum number of validations when no agreement achieved. The fourth row reports the number of validations performed by the users. The average elapsed time that users spent to validate a mapping is reported in the last row. Observe that, the performance without validation, in the first column, is 50%, since 50% of the sample dataset (Table 7.1) are decidable mappings, i.e., the candidate matches include the correct match that is ranked as TopOne.

The defined relationships are split as follows. In the BRAVE validation; 11 of type specific relation, 8 of type general relation, and 149 of type equivalent relation; in the CAUTIOUS validation: 16 of type specific relation, 12 of type general relation, and 171 of type equivalent relation. Based on the minimum consensus agreement approach users effort is reduced by 5.8% and 15.4% in the BRAVE and CAUTIOUS validations restrictively, w.r.t the maximum number of the required validations.

An important observation, is that, users have not reached an agreement in the High-difficulty validation in most cases. This is due to the fact that the available evidence, even for the users, are not sufficient to decide and select the correct relation. If definitions or examples (sense tagged sentences) are available for the source concepts, i.e., any further contextual knowledge, it would be easier for the users to select the correct relation. For instance, in most of the High-difficulty validations users accessed the online glossary aiming to find more evidence, however, the glossary provides all the possible definitions (senses) of the word without disambiguating its sense. While information provided about the candidate matches (*Matches Info*) seems to be sufficient for the users, few of them accessed the online WordNet in order to navigate the wordnet hierarchic. In fact, this confirms the usefulness of the classification method, defined in Chapter 5, and the correctness of estimating the difficulty of the mapping selection tasks based on the available evidence.

The average elapsed time in the CAUTIOUS validation is less than the time in the BRAVE validation; one reason might be due to the increase of users awareness of the system.

7.6 Conclusions

In this Chapter, I presented a suggestion-based cross-lingual mapping system, called Interactive Cross-Lingual Mapping (ICLM), which supports users with quality mappings by leveraging translation evidence and lexical characteristics using the LSOA. ICLM reduces the users effort by distributing the mapping tasks to a different number of users based on an estimated difficulty of these mappings, and accordingly collects users feedback in more efficient way, in contrast to pure crowdsourcing models where tasks are equally assigned to a fixed number of users. A user study is conducted to evaluate ICLM's strategies in estimating and distributing the validation tasks. The experimental results provide evidence that the estimated difficulties to a large extent are precise, and the classification method used to classify these task is useful.

As a future direction, I plan to investigate further strategies to distribute the validation tasks over users. For instance, I would like to investigate an active learning model presented in [26]. Another interesting direction would be to consider more languages and incorporate more users. In addition, to learn from users behavior in order to reconfigure the difficulty estimation is another interesting direction to explore. Moreover, an in-depth analyze w.r.t each concept category should be also considered.

Chapter 8

Conclusion and Future Work

This chapter summarizes the main contributions of this thesis in Section 8.1, followed by an outlook on further research directions for future work in Section 8.2.

8.1 Summary of Contributions

This thesis has analyzed at a large-scale the lexical evidence of automatic translations, which is efficiently used by an automatic cross-lingual mapping method. This method decides the best mapping, among a set of candidate matches, using a novel translation-based similarity measure (TSM) that exploits the translations evidence, and efficiently selects the appropriate mapping using a local similarity optimization algorithm (LSOA), which leverages the lexical characteristics of concepts. This thesis has investigated at a large-scale the lexical characteristics of concepts in lexical ontologies in different families of languages; a novel classification method for concepts has been defined. An interactive mapping approach (ICLM) takes advantage from this classification, which estimates the difficulty of the mapping tasks and accordingly estimates the number of users that are required to validate the mappings, this speeds up the validation process and improves automatically generated alignments.

This thesis has investigated the limits and upper bounds of using lexical evidence for processing cross-lingual mapping tasks, by exploiting evidence collected from automatic translations and lexical characteristic of the concepts. Nevertheless, the impact of structural evidence on the mapping process is also addressed in this thesis as a minor contribution.

The research presented in this thesis has led to five peer-reviewed scientific publications (Listed in Appendix A), which also constitute the scientific accomplishment of the author.

This thesis has proposed innovative techniques for efficiently performing very large cross-lingual mapping tasks, which can be summarized as follows.

Classification-based interpretation for cross-lingual mappings. I have addressed this issue by adapting a classification-based semantics for cross-lingual ontology mapping, taking into account the lexicalization of concepts. I defined the classification task as disambiguation task, which I refer to as cross-lingual word sense disambiguation task (CL-WSD), namely, the classification of a word in a given sentence in one language as occurrence of a word sense lexicalized in another language. To the best of my knowledge, this is the first attempt to provide a formal interpretation of cross-lingual mappings. This formal semantic cross-lingual mapping for lexically founded ontologies would allow to define a set of inference rules to derive semantic relations from a set of existing mappings. This would be helpful in scenarios where unstructured resource (e.g., a dictionary) is mapped to a structured resource (e.g, lexical ontology), hence a novel relations can be derived from relations in the target resource.

Classification of concepts. I have classified concepts in lexical ontologies into different categories, based on different characteristics: word ambiguity (e.g., monosemous vs polysemous), number of synonyms (e.g., synonymful vs synonymless), and concept specificity (e.g., leaves vs intermediate concepts). This classification is useful in evaluating and comparing the performance of different automatic translations on the cross-lingual mapping tasks in different ways. First, to evaluate and compare the performance of mapping methods based on different category of concepts in contrast to a global evaluation. Second, to evaluate the effectiveness of different translation resources, where the upper-bounds (coverage) and the correctness of translations are measured based on the different categories. Third, to estimate the difficulty of the mapping selection tasks based on the category of concepts involved in the corespondent mapping task.

Effectiveness of automatic translations on cross-lingual mapping tasks. In spite of the quality of translations that machine translation tools provide, where noise translations appear, they are largely available for different pair of languages, including resource-poor languages. Based on a large-scale study, I observed that machine translation tools can provide a sufficient evidence to support the decision in generating quality mappings.

Efficient cross-lingual mapping method for very large lexical ontologies. I have provided an efficient disambiguation technique based on a local similarity optimization algorithm (LSOA). The algorithm takes advantage of translations evidence and lexical characteristics of mapped concepts in order to select the appropriate mapping among a set of candidate matches. The translations evidence is used to measure the degree of similarity between the concepts using the translation-based similarity measure (TSM), whereas a disambiguation graph exploits the lexical characteristic of the concepts; this graph is used to reduce the mapping space, and hence to efficiently select the best mappings using an optimization algorithm. Thus, this method would support the process of mapping very large lexical resources and provides quality mappings.

Interactive mapping. Automatic mapping methods are neither correct nor complete when compared against gold standards. Based on the available evidence, some mappings may still be hard to decide upon using a fully automatic approach (e.g., when ties occur), thus the mapping process also requires feedback provided by users, especially in real-world scenarios. To this end, a Web tool called ICLM has been implemented to collect users feedback. ICLM tries to reduce users effort by distributing the mapping tasks to a different number of users based on an estimated difficulty of the mapping tasks. ICLM estimates the difficulty of the mapping tasks based on lexical characteristics of concepts under evaluation, where the confidence of candidate matches is measured based on TSM and LSOA.

8.2 Future Work

The research shown in this thesis is a step on improving cross-lingual mapping quality through the use of translations evidence to support the decision of selecting the appropriate mappings. The work presented in this thesis opens up several research opportunities for future work, discussed next.

Evaluation:

Firstly, additional evaluation experiments with more ontology pairs involving specific domains (e.g., OAEI multifarm dataset [77]) and natural languages (e.g., wordnets not considered in this thesis [14]) will give further insight into the use of the proposed mapping method in the process of cross-lingual mapping.

Secondly, more case studies involving more users should be developed to evaluate in depth the usefulness and usability of the interactive mapping approach ICLM. A goal-oriented approach for ontology mapping is one direction to follow in the evaluation process. For example, ICLM can be evaluated be a user-centric approach and focus on how well a particular validation task is performed with the assistance of ICLM [87], or to consider an end-to-end evaluation approach whereby evaluations are carried out on the performance of the applications (e.g., cross-lingual information retrieval [83, 23], cross-lingual word sense disambiguation [67], Web tables annotation [117, 82]) that consume the mappings produced by the mapping methods [54].

Implementation:
Firstly, incorporating structural evidence has shown to improve the quality of the mappings. However, different structural-based matching method should be developed (or, used) to evaluate the performance of the generated mappings, e.g., [78, 29, 104, 90]. In addition, the way to incorporate the structural information with the proposed mapping method should be further investigated, i.e, different configurations can be considered to incorporate these information. Thus, further experiments are necessary in this direction.

Secondly, the use of users feedback in ICLM can be expanded to assist the generation of higher quality mappings. For instance, user behaviors may be used to assist the estimation of the difficulty level of the validation tasks. Moreover, an active learning model with pay-as-you-go model, e.g., similar to the work presented in [26], can be used to consider the users explicit feedback to generate more reliable mappings, which recompute the alignment after every user feedback.

Thirdly, explores mappings uncertainty to establish semantic relations between concepts in unstructured resources. This would be addressed though applying inference role on top of the mapping method in order to drive and assist the semantic relations in these resources [47, 64].

Fourthly, the local similarity optimization algorithm (LSOA) should be further investigated to enhance its performance. For example, to extend the Hungarian algorithm so that it does not stop after the generation of the (first) best mapping but continues to the generation of the best $top - k$ mappings, similar to work presented in [12]. Further, to advance the way that the algorithm merges the locally generated mappings in contrast to the greedy approach that has been followed in the thesis.

Lastly, further development on ICLM should be considered to facilitate the validation process. For example, enable users to visualize the generated mappings and relations [66]. Moreover, providing open-source API to help the advancement of this field.

Appendix A

Published Work

Parts of the work presented in this thesis have been published in international journals and the proceedings of international conferences and refereed workshops. Publications relating to this work are listed below:

- International Journals
  - [5] Abu Helou, M., Palmonari, M., and Jarrar, M. Effectiveness of automatic translations for cross-lingual ontology mapping. *J. Artif. Intell. Res. (JAIR). Special Track on Cross-language Algorithms and Applications 55*, 165–208, (2016).
- International Conferences
  - [6] Abu Helou, M., Palmonari, M., Jarrar, M., and Fellbaum, C. Towards building linguistic ontology via cross-language matching. In *Proceedings of the 7th International Conference on Global WordNet* (2014).
  - [4] Abu Helou, M., and Palmonari, M. Upper bound for cross-lingual concept mapping with external translation resources. In *20th Natural Language Processing and Information Systems (NLDB 2015)*, vol. 9103 of *Lecture Notes in Computer Science*. Springer International Publishing, pp. 424–431. (2015).
  - [3] Abu Helou, M., and Palmonari, M. Cross-lingual lexical matching with word translation and local similarity optimization. In

Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2015, Vienna, Austria, September. (2015).

- Refereed Workshops

  – [1] Abu Helou, M. Towards constructing linguistic ontologies: Mapping framework and preliminary experimental analysis. In *Proceedings of the Second Doctoral Workshop in Artificial Intelligence (DWAI 2014) An official workshop of the 13th Symposium of the Italian Association for Artificial Intelligence "Artificial Intelligence for Society and Economy" (AI*IA 2014), Pisa, Italy.* (2014).

- Technical Reports

  – [2] Abu Helou, M., Jarrar, M., Palmonari, M., Salhi, A., Hicks, A., Bortoli, S., Roche, C., Fellbaum, C., Bouquet, P., and Yahya, A. Arabization and multilingual knowledge sharing-intermediate report on research setup. In *SIERA Project 2.1 Deliverable* (May 2013).

  – [62] Jarrar, M., Yahya, A., Salhi, A., Abu Helou, M., Sayrafi, B., Arar, M., Daher, J., Hicks, A., Fellbaum, C., Bortoli, S., Bouquet, P., Costa, R., Roche, C., and Palmonari, M. Arabization and multilingual knowledge sharing- final report on research setup. In *SIERA Project 2.3 Deliverable* (September 2014).

Bibliography

[1] Abu Helou, M. Towards constructing linguistic ontologies: Mapping framework and preliminary experimental analysis. In *Proceedings of the 2nd Doctoral Workshop in Artificial Intelligence (DWAI 2014) An official workshop of the 13th Symposium of the Italian Association for Artificial Intelligence (AI*IA 2014), Pisa, Italy* (2014).

[2] Abu Helou, M., Jarrar, M., Palmonari, M., Salhi, A., Hicks, A., Bortoli, S., Roche, C., Fellbaum, C., Bouquet, P., and Yahya, A. Arabization and multilingual knowledge sharing- intermediate report on research setup. In *SIERA Project 2.1 Deliverable* (May 2013).

[3] Abu Helou, M., and Palmonari, M. Cross-lingual lexical matching with word translation and local similarity optimization. In *Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2015, Vienna, Austria, September* (2015).

[4] Abu Helou, M., and Palmonari, M. Upper bound for cross-lingual concept mapping with external translation resources. In *20th Natural Language Processing and Information Systems (NLDB 2015)*, vol. 9103 of *Lecture Notes in Computer Science*. Springer International Publishing, 2015.

[5] Abu Helou, M., Palmonari, M., and Jarrar, M. Effectiveness of automatic translations for cross-lingual ontology mapping. *J. Artif. Intell. Res. (JAIR) Special Track on Cross-language Algorithms and Applications 55* (2016), 165–208.

[6] Abu Helou, M., Palmonari, M., Jarrar, M., and Fellbaum, C. Towards building linguistic ontology via cross-language matching. In *Proceedings of the 7th International Conference on Global WordNet* (2014).

[7] Al-Kabi, M. N., Hailat, T. M., Al-Shawakfa, E. M., and Alsmadi, I. M. Evaluating English to Arabic Machine Translation Using BLEU. *International Journal of Advanced Computer Science and Applications(IJACSA) 4*, 1 (2013).

[8] Apidianaki, M. Data-driven semantic analysis for multilingual wsd and lexical selection in translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (Stroudsburg, PA, USA, 2009), EACL '09, Association for Computational Linguistics, pp. 77–85.

[9] Atencia, M., Borgida, A., Euzenat, J., Ghidini, C., and Serafini, L. A formal semantics for weighted ontology mappings. In *International Semantic Web Conference (1)* (2012), pp. 17–33.

[10] Bellahsene, Z., Bonifati, A., and Rahm, E., Eds. *Schema Matching and Mapping*. Springer, 2011.

[11] Bentivogli, L., Forner, P., Magnini, B., and Pianta, E. Revising the wordnet domains hierarchy: Semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources* (Stroudsburg, PA, USA, 2004), MLR '04, Association for Computational Linguistics, pp. 101–108.

[12] Bergamaschi, S., Domnori, E., Guerra, F., Lado, R. T., and Velegrakis, Y. Keyword search over relational databases: a metadata approach. In *SIGMOD Conference* (2011), T. K. Sellis, R. J. Miller, A. Kementsietsidis, and Y. Velegrakis, Eds., ACM.

[13] Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. *Scientific American 284*, 5 (may 2001), 34–43.

[14] Bond, F., and Foster, R. Linking and extending an open multilingual wordnet. In *ACL (1)* (2013), The Association for Computer Linguistics, pp. 1352–1362.

[15] Bouma, G. Cross-lingual ontology alignment using eurowordnet and wikipedia. In *LREC* (2010), N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, Eds., European Language Resources Association.

[16] Bouquet, P., Ehrig, M., Euzenat, J., Franconi, E., Hitzler, P., (u. Karlsruhe, M. E., (inria, J. E., (fu, E. F., (u. Karlsruhe, P. H., (tu Dresden, M. K., (fu Bolzano, S. T., Fensel, C. P. D., and Leger, C. P. A. D2.2.1 specification of a common framework for characterizing alignment,technical report,knowledge web, 2004.

[17] Bouquet, P., Serafini, L., and Zanobini, M. Semantic coordination in systems of autonomous agents: the approach and an implementation. In *WOA* (2003), pp. 179–186.

[18] Bourgeois, F., and Lassalle, J.-C. An extension of the munkres algorithm for the assignment problem to rectangular matrices. *Commun. ACM 14*, 12 (Dec. 1971).

[19] Boyd-Graber, J., , Osherson, and Schapire, R. Adding dense, weighted connections to WordNet connections to wordnet. In *Proceedings of the Third Global WordNet Meeting* (2006).

[20] Carruthers, P. The cognitive functions of language. *Behavioral and Brain Sciences 25*, 6 (2002), 657–674.

[21] Choi, N., Song, I.-Y., and Han, H. A survey on ontology mapping. *SIGMOD Rec. 35*, 3 (Sept. 2006), 34–41.

[22] Cimiano, P., Montiel-Ponsoda, E., Buitelaar, P., Espinoza, M., and Gmez-Prez, A. A note on ontology localization. *Applied Ontology 5*, 2 (2010), 127–137.

[23] Cimiano, P., Schultz, A., Sizov, S., Sorg, P., and Staab, S. Explicit versus latent concept models for cross-language information retrieval. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence* (San Francisco, CA, USA, 2009), IJCAI'09, Morgan Kaufmann Publishers Inc., pp. 1513–1518.

[24] Cohen, W. W., Ravikumar, P. D., and Fienberg, S. E. A comparison of string distance metrics for name-matching tasks. In *IIWeb* (2003), pp. 73–78.

[25] Costa-juss, M. R., Farrs, M., Mario, J. B., and Fonollosa, J. A. R. Study and comparison of rule-based and statistical catalan-spanish

machine translation systems. *Computing and informatics 31*, 2 (2012), 245–270.

[26] Cruz, I., Loprete, F., Palmonari, M., Stroe, C., and Taheri, A. Pay-as-you-go multi-user feedback model for ontology matching. In *Knowledge Engineering and Knowledge Management*, K. Janowicz, S. Schlobach, P. Lambrix, and E. Hyvnen, Eds., vol. 8876 of *Lecture Notes in Computer Science*. Springer International Publishing, 2014, pp. 80–96.

[27] Cruz, I. F., Antonelli, F. P., and Stroe, C. Agreementmaker: Efficient matching for large real-world schemas and ontologies. *PVLDB* (2009).

[28] Cruz, I. F., Palmonari, M., Caimi, F., and Stroe, C. Building linked ontologies with high precision using subclass mapping discovery. *Artif. Intell. Rev. 40*, 2 (Aug. 2013), 127–145.

[29] Cruz, I. F., and Sunna, W. Structural alignment methods with applications to geospatial ontologies. In *Transactions in GIS, special issue on Semantic Similarity Measurement and Geospatial Applications* (2008).

[30] de Melo, G., and Weikum, G. Towards a universal wordnet by learning from combined evidence. In *CIKM* (2009), D. W.-L. Cheung, I.-Y. Song, W. W. Chu, X. Hu, and J. J. Lin, Eds., ACM, pp. 513–522.

[31] de Melo, G., and Weikum, G. Constructing and utilizing wordnets using statistical methods. *Language Resources and Evaluation 46*, 2 (2012), 287–311.

[32] Djeddi, W. E., and Khadir, M. T. Xmap++: Results for oaei 2014. In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October* (2014), vol. 20, pp. 163–169.

[33] Doan, A., and Halevy, A. Y. Semantic integration research in the database community: A brief survey. *AI Magazine 26*, 1 (2005), 83–94.

[34] dos Santos, C. T., Quaresma, P., and Vieira, R. An api for multilingual ontology matching. In *LREC* (2010).

[35] Eidoon, Z., Yazdani, N., and Oroumchian, F. A vector based method of ontology matching. In *Semantics, Knowledge and Grid, Third International Conference on* (2007), pp. 378–381.

[36] Els, L., and Vronique, H. Semeval-2010 task 3: Cross-lingual word sense disambiguation. Proceedings of the 5th International Workshop on Semantic Evaluation, Association for Computational Linguistics.

[37] Euzenat, J., and Shvaiko, P. *Ontology matching.* Springer, 2007.

[38] Faria, D., Martins, C., Nanavaty, A., Taheri, A., Pesquita, C., Santos, E., Cruz, I. F., and Couto, F. M. Agreementmakerlight results for oaei 2014. In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October* (2014), vol. 20, pp. 126–134.

[39] Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., and Couto, F. M. The agreementmakerlight ontology matching system. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences* (2013), Springer, pp. 527–541.

[40] Fellbaum, C., Ed. *WordNet An Electronic Lexical Database.* The MIT Press, Cambridge, MA ; London, May 1998.

[41] Fiŝer, D. Leveraging parallel corpora and existing wordnets for automatic construction of the slovene wordnet. In *LTC* (2007), Z. Vetulani and H. Uszkoreit, Eds., vol. 5603 of *Lecture Notes in Computer Science*, Springer, pp. 359–368.

[42] Fodor, J. *The Language of Thought.* Cambridge, MA: Harvard University Press, 1975.

[43] Fu, B., Brennan, R., and O'Sullivan, D. Cross-lingual ontology mapping - an investigation of the impact of machine translation. In *ASWC* (2009), pp. 1–15.

[44] Fu, B., Brennan, R., and O'Sullivan, D. A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes. *J. Web Semantic 15* (2012), 15–36.

[45] Gale, W. A., Church, K. W., and Yarowsky, D. Using bilingual materials to develop word sense disambiguation methods. In *In Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation* (1992).

[46] Garcia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., and McCrae, J. Challenges for the multilingual web of data. *Web Semantics: Science, Services and Agents on the World Wide Web 11*, 0 (2012).

[47] Gómez-Pérez, A., and Manzano-Macho, D. An overview of methods and tools for ontology learning from texts. *The Knowledge Engineering Review 19*, 3 (2004), 187–212.

[48] Gómez-Pérez, A., Vila-Suero, D., Montiel-Ponsoda, E., Gracia, J., and Aguado-de Cea, G. Guidelines for multilingual linked data. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics* (New York, NY, USA, 2013), WIMS '13, ACM, pp. 3:1–3:12.

[49] Gonzalez-Agirre, A., Laparra, E., and Rigau, G. Multilingual central repository version 3.0. In *LREC* (2012), N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA), pp. 2525–2529.

[50] Gruber, T. R. A translation approach to portable ontology specifications. *Knowl. Acquis. 5*, 2 (June 1993), 199–220.

[51] Hertling, S., and Paulheim, H. Wikimatch - using wikipedia for ontology matching. In *OM* (2012), P. Shvaiko, J. Euzenat, A. Kementsietsidis, M. Mao, N. F. Noy, and H. Stuckenschmidt, Eds., vol. 946 of *CEUR Workshop Proceedings*, CEUR-WS.org.

[52] Hirst, G. Ontology and the lexicon. In *Handbook on Ontologies and Information Systems* (2004), eds. S. Staab and R. Studer, Eds., Heidelberg, Springer.

[53] Hirst, G. Overcoming linguistic barriers to the multilingual semantic web. In *Towards the Multilingual Semantic Web*, P. Buitelaar and P. Cimiano, Eds. Springer Berlin Heidelberg, 2014, pp. 3–14.

[54] Hollink, L., Assem, M. V., Wang, S., and Isaac, A. Two variations on ontology alignment evaluation: Methodological issues. In *In Proceedings of the 5th European Semantic Web Conference, ESWC* (2008).

[55] Horrocks, I. Ontologies and the semantic web. *Communications of the ACM 51*, 12 (2008), 58–67.

[56] Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., and Weischede, R. Ontonotes: the 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL. Companion Volume, Short Papers on XX, NAACL 06. Association for Computational Linguistics (ACL)* (Morristown, NJ, USA, 2006).

[57] Hovy, E., Navigli, R., and Ponzetto, S. P. Collaboratively built semi-structured content and artificial intelligence: The story so far. *Artificial Intelligence* (2012).

[58] Ide, N., Erjavec, T., and Tufis, D. Sense discrimination with parallel corpora. In *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions - Volume 8* (Stroudsburg, PA, USA, 2002), WSD '02, Association for Computational Linguistics, pp. 61–66.

[59] Isaac, A., Van Der Meij, L., Schlobach, S., and Wang, S. An empirical study of instance-based ontology matching. In *Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference* (Berlin, Heidelberg, 2007), ISWC'07/ASWC'07, Springer-Verlag, pp. 253–266.

[60] Jarrar, M. Building a formal arabic ontology (invited paper). In *Proceedings of the Experts Meeting On Arabic Ontologies And Semantic Networks. Alecso, Arab League. 26-28 July.Tunis* (2011).

[61] Jarrar, M., Rishmawi, R., and Olwan, H. Top levels of the arabic ontology. technical report. sina institute, birzeit university, 2013.

[62] Jarrar, M., Yahya, A., Salhi, A., Abu Helou, M., Sayrafi, B., Arar, M., Daher, J., Hicks, A., Fellbaum, C., Bortoli, S., Bouquet, P., Costa, R., Roche, C., and Palmonari, M. Arabization and multilingual knowledge sharing- final report on research setup. In *SIERA Project 2.3 Deliverable* (2014).

[63] Jiménez-Ruiz, E., and Grau, B. C. Logmap: Logic-based and scalable ontology matching. In *The Semantic Web–ISWC 2011*. Springer, 2011, pp. 273–288.

[64] Jiménez-Ruiz, E., Grau, B. C., Xia, W., Solimando, A., Chen, X., Cross, V., Gong, Y., Zhang, S., and Chennai-Thiagarajan, A.

Logmap family results for oaei 2014? In *Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October* (2014), vol. 20, pp. 126–134.

[65] Kuhn, H. W. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly 2*, 1–2 (March 1955).

[66] Lanzenberger, M., Sampson, J., and Rester, M. Ontology visualization: Tools and techniques for visual representation of semi-structured meta-data. *Journal of Universal Computer Science 16*, 7 (2008).

[67] Lefever, E., and Hoste, V. Semeval-2013 task 10: Cross-lingual word sense disambiguation. *Proc. of SemEval* (2013), 158–166.

[68] Liang, A. C., and Sini, M. Mapping agrovoc and the chinese agricultural thesaurus: Definitions, tools, procedures. *The New Review of Hypermedia and Multimedia 12*, 1 (2006), 51–62.

[69] Lin, F., Butters, J., Sandkuhl, K., and Ciravegna, F. Context-based ontology matching: Concept and application cases. In *10th IEEE International Conference on Computer and Information Technology, CIT 2010, Bradford, West Yorkshire, UK, June 29-July 1, 2010* (2010), pp. 1292–1298.

[70] Lin, F., and Krizhanovsky, A. Multilingual ontology matching based on wiktionary data accessible via sparql endpoint. *CoRR abs/1109.0732* (2011).

[71] Lin, F., and Sandkuhl, K. A survey of exploiting wordnet in ontology matching. In *IFIP AI* (2008), pp. 341–350.

[72] Margolis, E., and Laurence, S. Concepts. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., spring 2014 ed. 2014.

[73] Mariòo, J. B., Banchs, R. E., Crego, J. M., de Gispert, A., Lambert, P., Fonollosa, J. A. R., and Costa-jussà, M. R. N-gram-based machine translation. *Comput. Linguist. 32*, 4 (Dec. 2006), 527–549.

[74] McCarley, J. S. Should we translate the documents or the queries in cross-language information retrieval? In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational*

Linguistics (Stroudsburg, PA, USA, 1999), ACL '99, Association for Computational Linguistics, pp. 208–214.

[75] McCarthy, D., and Navigli, R. The english lexical substitution task. *Language Resources and Evaluation 43*, 2 (2009), 139–159.

[76] McCrae, J., Spohr, D., and Cimiano, P. Linking lexical resources and ontologies on the semantic web with lemon. In *Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web: Research and Applications - Volume Part I* (Berlin, Heidelberg, 2011), ESWC'11, Springer-Verlag, pp. 245–259.

[77] Meilicke, C., Garcia-Castro, R., Freitas, F., van Hage, W. R., Montiel-Ponsoda, E., de Azevedo, R. R., Stuckenschmidt, H., Svb-Zamazal, O., Svtek, V., Tamilin, A., dos Santos, C. T., and Wang, S. Multifarm: A benchmark for multilingual ontology matching. *J. Web Sem. 15* (2012), 62–68.

[78] Melnik, S., Garcia-Molina, H., and Rahm, E. Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the 18th International Conference on Data Engineering* (Washington, DC, USA, 2002), ICDE '02, IEEE.

[79] Miller, G. A. Wordnet: A lexical database for english. *Commun. ACM 38*, 11 (Nov. 1995), 39–41.

[80] Miller, G. A., Leacock, C., Tengi, R., and Bunker, R. T. A semantic concordance. In *Proceedings of the workshop on Human Language Technology* (Stroudsburg, PA, USA, 1993), HLT '93, Association for Computational Linguistics, pp. 303–308.

[81] Mochól, M. *The methodology for finding suitable ontology matching approaches.* PhD thesis, Freie Universitat Berlin, Germany, 2009.

[82] Mulwad, V., Finin, T., and Joshi, A. Semantic message passing for generating linked data from tables. In *International Semantic Web Conference (1)* (2013), H. Alani, L. Kagal, A. Fokoue, P. T. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. F. Noy, C. Welty, and K. Janowicz, Eds., vol. 8218 of *Lecture Notes in Computer Science*, Springer, pp. 363–378.

[83] Narducci, F., Palmonari, M., and Semeraro, G. Cross-language semantic matching for discovering links to e-gov services in the lod cloud. In *KNOW@LOD* (2013), J. Vlker, H. Paulheim, J. Lehmann, M. Niepert, and H. Sack, Eds., vol. 992 of *CEUR Workshop Proceedings*, CEUR-WS.org, pp. 21–32.

[84] Nastase, V., Strube, M., Boerschinger, B., Zirn, C., and Elghafari, A. Wikinet: A very large scale multi-lingual concept network. In *LREC* (2010), N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, Eds., European Language Resources Association.

[85] Navigli, R. Word sense disambiguation: A survey. *ACM Comput. Surv. 41*, 2 (Feb. 2009), 10:1–10:69.

[86] Navigli, R., and Ponzetto, S. P. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence 193*, 0 (2012), 217 – 250.

[87] Noy, N. F., and Musen, M. A. Evaluating ontology-mapping tools: Requirements and experience. In *In Workshop on Evaluation of Ontology Tools at EKAW, Spain* (Siguenza, Spain, 2002).

[88] Obrst, L. Ontologies for semantically interoperable systems. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management* (New York, NY, USA, 2003), CIKM '03, ACM, pp. 366–369.

[89] Oliver, A., and Climent, S. Building wordnets by machine translation of sense tagged corpora. In *Proceedings of the 6th International Conference on Global WordNet* (2012).

[90] Otero-Cerdeira, L., Rodrguez-Martnez, F. J., and Gmez-Rodrguez, A. Ontology matching: A literature review. *Expert Systems with Applications 42*, 2 (2015), 949 – 971.

[91] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (Stroudsburg, PA, USA, 2002), ACL '02, Association for Computational Linguistics, pp. 311–318.

[92] Pazienza, M., and Stellato, A. An open and scalable framework for enriching ontologies with natural language content. In *Advances in Applied Artificial Intelligence*, M. Ali and R. Dapoigny, Eds., vol. 4031 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, pp. 990–999.

[93] Pianta, E., Bentivogli, L., and Girardi, C. Multiwordnet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet* (January 2002).

[94] Pinker, S. *The Language Instinct.* Harper Perennial Modern Classics, New York, 1994.

[95] Po, L., and Sorrentino, S. Automatic generation of probabilistic relationships for improving schema matching. *Information Systems 36*, 2 (2011), 192 – 208. Special Issue: Semantic Integration of Data, Multimedia, and Services.

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

[97] Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Mart, M. A., Black, W., Elkateb, S., Kirk, J., Vossen, P., and Fellbaum, C. Arabic wordnet: Current state and future extensions. In *Proceedings of the Forth International Conference on Global WordNet* (2008).

[98] Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. Multiword expressions: A pain in the neck for nlp. In *Computational Linguistics and Intelligent Text Processing*, vol. 2276 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2002, pp. 1–15.

[99] Sagot, B., and Fišer, D. Building a free french wordnet from multilingual resources. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC)* (2008).

[100] Sarasua, C., Simperl, E., and Noy, N. F. Crowdmap: Crowdsourcing ontology alignment with microtasks. In *Proceedings of the 11th International Semantic Web Conference (ISWC2012)* (2012).

[101] Saveski, M., and Trajkovski, I. Automatic construction of wordnets by using machine translation and language modeling. In *In Proceedings of Seventh Language Technologies Conference, 13th International Multiconference Information Society, volume C* (2010).

[102] Shvaiko, P. *Iterative schema-based semantic matching.* PhD thesis, University of Trento, Italy, 2006.

[103] Shvaiko, P., and Euzenat, J. Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng. 25*, 1 (2013), 158–176.

[104] Shvaiko, P., Euzenat, J., Mao, M., Jiménez-Ruiz, E., Li, J., and Ngonga, A., Eds. *Proceedings of the 9th International Workshop on Ontology Matching collocated with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Trentino, Italy, October 20, 2014* (2014), vol. 1317 of *CEUR Workshop Proceedings*, CEUR-WS.org.

[105] Sorrentino, S., Bergamaschi, S., Gawinecki, M., and Po, L. Schema label normalization for improving schema matching. *Data & Knowledge Engineering 69*, 12 (2010), 1254 – 1273. Special issue on 28th International Conference on Conceptual Modeling (ER 2009).

[106] Spelke, S. E. What makes us smart? core knowledge and natural language. In *Language in Mind: Advances in the Study of Language and Thought.* Mit Press, 2003, pp. 277–311.

[107] Spohr, D., Hollink, L., and Cimiano, P. A machine learning approach to multilingual and cross-lingual ontology matching. In *Proceedings of the 10th International Conference on The Semantic Web - Volume Part I* (Berlin, Heidelberg, 2011), ISWC'11, Springer-Verlag, pp. 665–680.

[108] Spohr, D., Hollink, L., and Cimiano, P. A machine learning approach to multilingual and cross-lingual ontology matching. In *International Semantic Web Conference (1)* (2011), pp. 665–680.

[109] Suchanek, F. M., Kasneci, G., and Weikum, G. Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web 6*, 3 (2008), 203–217.

[110] Tomaž, E., and Fišer, D. Building slovene wordnet. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)* (Genoa, Italy, 2006).

[111] Trojahn, C., Fu, B., Zamazal, O., and Ritze, D. *State-of-the-art in Multilingual and Cross-Lingual Ontology Matching.* Springer, 2014.

[112] Tufis, D., Cristea, D., and Stamou, S. Balkanet: Aims, methods, results and perspectives. a general overview. *In: D. Tufi (ed): Special Issue on BalkaNet. Romanian Journal on Science and Technology of Information* (2004).

[113] VAN DIJK, T. A. Discourse context and cognition. discourse studies. *Discourse Studies 8* (2006), 159–177.

[114] Vincent, V. A. Macro- and micro-averaged evaluation measures [[basic draft]]. In *Technical report* (2013).

[115] Vossen, P. Eurowordnet: A multilingual database of autonomous and language-specific wordnets connected via an inter-lingualindex. *International Journal of Lexicography 17*, 2 (June 2004), 161–173.

[116] Vossen, P., Eneko, A., Francis, B., Wauter, B., Axel, H., Amanda, H., Shu-Kai, H., Hitoshi, I., Chu-Ren, H., Kyoko, K., Andrea, M., German, R., Francesco, R., Roxane, S., and Maurizio, T. *KYOTO: A Wiki for Establishing Semantic Interoperability for Knowledge Sharing Across Languages and Cultures.* Springer, 2010, pp. 265–294.

[117] Zhang, Z. Towards efficient and effective semantic table interpretation. In *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I* (2014), pp. 487–502.

[118] Zhuge, H., Xing, Y., and Shi, P. Resource space model, owl and database: Mapping and integration. *ACM Transactions on Internet Technology (TOIT) 8*, 4 (2008), 20.