A Generalized Constraint Approach to Bilingual Dictionary Induction for Low-Resource Language Families

ARBI HAZA NASUTION, Kyoto University and Universitas Islam Riau
YOHEI MURAKAMI and TORU ISHIDA, Kyoto University

The lack or absence of parallel and comparable corpora makes bilingual lexicon extraction a difficult task for low-resource languages. The pivot language and cognate recognition approaches have been proven useful for inducing bilingual lexicons for such languages. We propose constraint-based bilingual lexicon induction for closely related languages by extending constraints from the recent pivot-based induction technique and further enabling multiple symmetry assumption cycle to reach many more cognates in the transgraph. We further identify cognate synonyms to obtain many-to-many translation pairs. This article utilizes four datasets: one Austronesian low-resource language and three Indo-European high-resource languages. We use three constraint-based methods from our previous work, the Inverse Consultation method and translation pairs generated from Cartesian product of input dictionaries as baselines. We evaluate our result using the metrics of precision, recall, and F-score. Our customizable approach allows the user to conduct cross validation to predict the optimal hyperparameters (cognate threshold and cognate synonym threshold) with various combination of heuristics and number of symmetry assumption cycles to gain the highest F-score. Our proposed methods have statistically significant improvement of precision and F-score compared to our previous constraint-based methods. The results show that our method demonstrates the potential to complement other bilingual dictionary creation methods like word alignment models using parallel corpora for high-resource languages while well handling low-resource languages.

CCS Concepts: • **Computing methodologies** → **Language resources**; *Lexical semantics*;

Additional Key Words and Phrases: Constraint satisfaction problem, low-resource languages, closely-related languages, pivot-based bilingual lexicon induction, cognate recognition

ACM Reference format:
Arbi Haza Nasution, Yohei Murakami, and Toru Ishida. 2017. A Generalized Constraint Approach to Bilingual Dictionary Induction for Low-Resource Language Families. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 17, 2, Article 9 (November 2017), 29 pages.
https://doi.org/10.1145/3138815

This article is significantly extended from our previous work [20].
This research was partially supported by a Grant-in-Aid for Scientific Research (A) (17H00759, 2017-2020) and a Grant-in-Aid for Young Scientists (A) (17H04706, 2017-2020) from Japan Society for the Promotion of Science (JSPS). The first author was supported by Indonesia Endownment Fund for Education (LPDP).
Authors' addresses: A. H. Nasution, Department of Social Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan; email: arbi@ai.soc.i.kyoto-u.ac.jp and Department of Information Technology, Universitas Islam Riau, Jl. Kaharuddin Nasution 113, Pekanbaru, Riau 28284, Indonesia; email: arbi@eng.uir.ac.id; Y. Murakami, Unit of Design, Kyoto University, #506, KRP Bldg.9, 91 Chudoji Awata-cho, Shimogyo-ku, Kyoto 600-8815, Japan; email: yohei@i.kyoto-u.ac.jp; T. Ishida, Department of Social Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan; email: ishida@i.kyoto-u.ac.jp.

https://doi.org/10.1145/3138815

1 INTRODUCTION

Machine-readable bilingual dictionaries are very useful for information retrieval and natural language processing research but are usually unavailable for low-resource languages. Previous work on high-resource languages showed the effectiveness of parallel corpora [3, 8] and comparable corpora [7, 21] in inducing bilingual lexicons. Bilingual lexicon extraction is highly problematic for low-resource languages due to the paucity or outright omission of parallel and comparable corpora. The approaches of pivot language [31] and cognate recognition [15] have been proven useful in inducing bilingual lexicons for low-resource languages. Closely related languages share cognates that share most of the semantic or meaning of the lexicons [13]. Some linguistics studies [9, 34] show that the percentage of shared cognates, either related directly or via a synonym, constitutes a highly accurate linguistic distance measure based on mutual intelligibility, that is, the ability of speakers of one language to understand the other language. The higher the percentage of shared cognates between the languages, the lower the linguistic distance, the higher is the level of mutual intelligibility.

We recently introduced the promising approach of treating pivot-based bilingual lexicon induction for low-resource languages as an optimization problem [20] with cognate pair coexistence probability as a sole heuristic in the symmetry constraint. In this article, we propose generalized constraint-based bilingual lexicon induction for closely related languages by setting two steps to obtaining translation pair results. First, we identify one-to-one cognates by incorporating more constraints and heuristics to improve the quality of the translation result. We then identify the cognates' synonyms to obtain many-to-many translation pairs. In each step, we can obtain more cognate and cognate synonym pair candidates by iterating the n-cycle symmetry assumption until all possible translation pair candidates have been reached. We address the following research goals:

- *Creating many-to-many translation pairs between closely related languages*: Recognize cognates and cognate synonyms from direct and indirect connectivities via pivot word(s) by iterating the symmetry assumption cycle to improve the quality and quantity of the translation pair results.
- *Evaluating the generalized method performance*: We apply the Inverse Consultation method [31] and naive translation pairs generation from the Cartesian product of input dictionaries to all of our datasets and compare the results with those of our generalized methods using precision, recall, and F-score. We also conduct experiments with our previous constraint-based methods [20] with the same datasets and further conduct student's paired *t*-tests to show that our proposed methods have statistically significant improvement of precision and F-score. We conduct cross validation to predict the optimal hyperparameters (cognate threshold and cognate synonym threshold) to gain the highest F-score.

The rest of this article is organized as follows: In Section 2, we will briefly discuss related research on bilingual dictionary induction. Section 3 discusses closely related languages and existing methods in comparative linguistics. Section 4 details our strategy of recognizing cognate and cognate synonyms, core component for our proposal, which is described in Section 5. Section 6 introduces our experiment and the results. Finally, Section 7 concludes this article.

2 BILINGUAL DICTIONARY INDUCTION

An intermediate/pivot language approach has been applied in machine translation [32] and service computing [12] researches. The first work on bilingual lexicon induction to create bilingual dictionary between language A and language C via pivot language B is Inverse Consultation (IC) [31] by utilizing the structure of input dictionaries to measure the closeness of word meanings

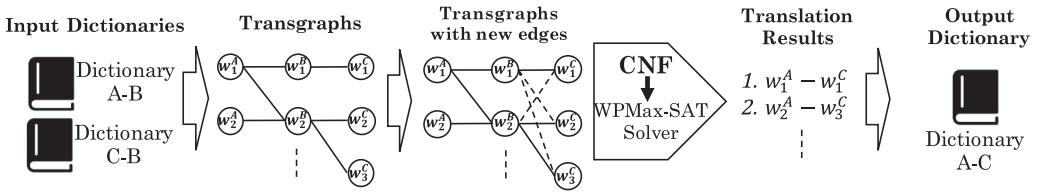


Fig. 1. One-to-one constraint approach to pivot-based bilingual dictionary induction.

and then use the results to prune erroneous translation pair candidates. The IC approach identifies equivalent candidates of language A words in language C by consulting dictionary A-B and dictionary B-C. These equivalent candidates will be looked up and compared in the inverse dictionary C-A. To analyze the method used to filter wrong translation pair candidates induced via the pivot-based approach, [24] explored distributional similarity measure (DS) in addition to IC. The analysis showed that IC depends on significant lexical variants in the dictionaries for each meaning in the pivot language, while DS depends on distributions or contexts across two corpora of the different languages. Their analysis also showed that the combination of IC and DS outperformed each used individually.

The pivot-based approach is very suitable for low-resource languages, especially when dictionaries are the only language resource required. Unfortunately, for some low-resource languages, it is often difficult to find machine-readable inverse dictionaries and corpora to filter the wrong translation pair candidates. Thus, we consider that the combination of IC and DS methods does not suit low-resource languages. To overcome this limitation, our team [37] proposed to treat pivot-based bilingual lexicon induction as an optimization problem. The assumption was that lexicons of closely related languages offer one-to-one mapping and share a significant number of cognates (words with similar spelling/form and meaning originating from the same root language). With this assumption, they developed a constraint optimization model to induce an Uyghur-Kazakh bilingual dictionary using Chinese language as the pivot, which means that Chinese words were used as intermediates to connect Uyghur words in an Uyghur-Chinese dictionary with Kazakh words in a Kazakh-Chinese dictionary. They used a graph whose vertices represent words and edges indicate shared meanings; they called this a transgraph following [29]. The steps in their approach are as follows: (1) use two bilingual dictionaries as input; (2) represent them as transgraphs where w_1^A and w_2^A are non-pivot words in language A, w_1^B and w_2^B are pivot words in language B, and w_1^C, w_2^C, and w_3^C are non-pivot words in language C; (3) add some new edges represented by dashed edges based on the one-to-one assumption; (4) formalize the problem into conjunctive normal form (CNF) and use the Weighted Partial MaxSAT (WPMaxSAT) solver [1] to return the optimized translation results; and (5) output the induced bilingual dictionary as the result. These steps are shown in Figure 1. The one-to-one approach depends only on semantic equivalence, one of the closely related language characteristics that permit the recognition of cognates between languages assuming that lexicons of closely related languages offer the one-to-one relation. If language A and C are closely related, then for any word in A there exists a unique word in C such that they have exactly the same meaning, and thus are symmetrically connected via pivot word(s). Such a pair is called a one-to-one pair. They realized that this assumption may be too strong for the general case, but they believed that it was reasonable for closely related languages like Turkic languages. They believe that their method works best for languages with high-similarity. They tried to improve the precision by utilizing multiple input dictionaries [36] while still applying the same one-to-one assumption. However, this assumption is too strong to be used for the induction of as many translation pairs as possible to offset resource paucity, because the few such pairs are yielded.

Table 1. Similarity Matrix of Top 10 Indonesian Ethnic Languages Ranked by Number of Speakers

Language	Indonesian	Malang	Yogyakarta	Old Javanese	Sundanese	Malay	Palembang Malay	Madurese	Minangkabau
Malang	23.46%								
Yogyakarta	27.29%	87.36%							
Old Javanese	24.09%	47.50%	52.18%						
Sundanese	39.43%	18.55%	22.43%	21.82%					
Malay	**85.10%**	20.53%	24.35%	21.36%	41.12%				
Palembang Malay	68.24%	33.97%	37.97%	31.85%	38.90%	73.23%			
Madurese	34.45%	17.63%	14.15%	15.18%	19.86%	34.16%	34.32%		
Minangkabau	**61.59%**	26.59%	29.63%	25.01%	30.81%	**61.66%**	63.60%	34.32%	
Buginese	31.21%	12.76%	16.85%	18.33%	24.80%	32.04%	31.00%	17.94%	32.00%

3 CLOSELY RELATED LANGUAGES

Historical linguistics is the scientific study of language change over time in term of sound, analogical, lexical, morphological, syntactic, and semantic information [4]. Comparative linguistics is a branch of historical linguistics that is concerned with language comparison to determine historical relatedness and to construct language families [13]. Many methods, techniques, and procedures have been utilized in investigating the potential distant genetic relationship of languages, including lexical comparison, sound correspondences, grammatical evidence, borrowing, semantic constraints, chance similarities, sound-meaning isomorphism, and so on [5]. The genetic relationship of languages is used to classify languages into language families. Closely related languages are those that came from the same origin or proto-language and belong to the same language family.

Automated Similarity Judgment Program (ASJP) was proposed in Reference [11] with the main goal of developing a database of Swadesh lists [30] for all of the world's languages from which lexical similarity or lexical distance matrix between languages can be obtained by comparing the word lists. We utilize ASJP to select our low-resource target languages for our first case study in this article. Indonesia has 707 low-resource ethnic languages [14] that are suitable as target languages in our study. There are three factors we consider in selecting the target languages: language similarity, input bilingual dictionary size, and number of speakers. To ensure that the induced bilingual dictionaries will be useful for many users, we listed the top 10 Indonesian ethnic languages ranked by the number of speakers. We then generated the language similarity matrix by utilizing ASJP as shown in Table 1. From this list, the biggest size machine-readable bilingual dictionaries are Minangkabau-Indonesian and Malay-Indonesian. After considering all those factors, we selected Malay, Minangkabau and Indonesian as our target languages for the low-resource languages case study.

Several machine translation studies focused on closely related languages [19, 25, 33]. In this research, the linguistic characteristics of the closely related languages play a vital role in improving quality of our method.

4 COGNATE AND COGNATE SYNONYM RECOGNITION

By utilizing linguistic information, we establish a strategy to obtain many-to-many translation pairs from a transgraph. The first step is to recognize one-to-one cognates in the transgraph that shares all their senses. Once a list of cognates is obtained, the next step is to recognize cognate synonyms in the transgraph; those that share part/all senses with the cognate and so are mutually connected to some/all pivot words. Those two steps are easy tasks when the input dictionaries have sense/meaning information as shown in Figure 2, where a cognate pair (w_1^A, w_1^C) share two senses, that is, s_1 and s_2 through pivot word w_1^B and a cognate pair (w_2^A, w_2^C) only share s_1 through pivot word w_1^B and w_2^B. Since for low-resource languages, a machine-readable bilingual dictionary with sense information is scarce, we regard connected words share at least one sense/meaning. Thus, we assume that non-pivot words that are symmetrically connected via pivot word(s) potentially share all their senses and so being a cognate.

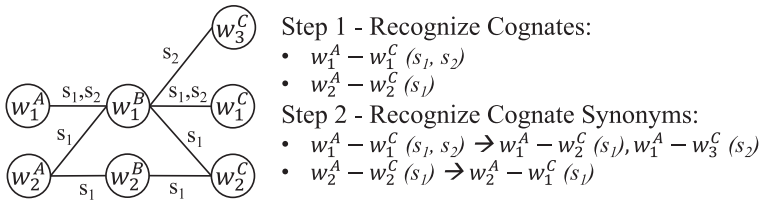


Fig. 2. Strategy to recognize cognates and cognate synonyms.

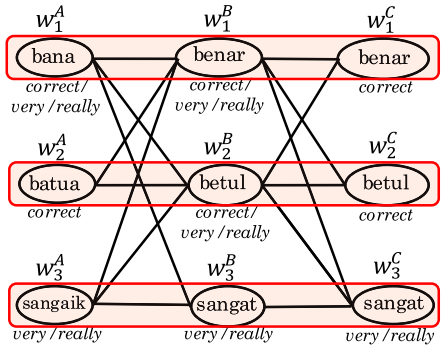
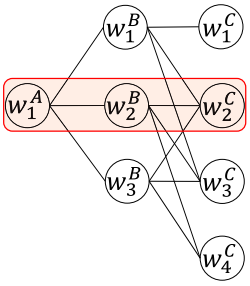


Fig. 3. Cognate and cognate synonym example.

Cognates are words with similar spelling/form and meaning that have a common etymological origin. For instance, the words *night* (English), *nuit* (French), *noche* (Spanish), *nacht* (German), and *nacht* (Dutch) have the same meaning, which is "night," and derived from the Proto-Indo-European **nók^{w}ts* with the same meaning of "night." Since most linguists believe that lexical comparison alone is not a good way to recognize cognates [4], we want to utilize a more general and basic characteristic of closely related languages, which is as follows: A cognate pair mostly maintain the semantic or meaning of the lexicons. Even though there is a possibility of a change in one of the meanings of a word in a language, within the families where the languages are known to be closely related, the possibility of a change is smaller. Since our approach targets the closely related languages, it is safe to make the following assumption based on the semantic characteristic of closely related languages: *Given a pair of words, w_i^A of language A and w_k^C of language C, if they are cognates, then they share all of their senses/meanings and are symmetrically connected through pivot word(s) from language B.* We call this the symmetry assumption. Unfortunately, in some cases, symmetry assumption is inadequate to eliminate wrong cognate from the cognate pair candidates when a pivot-word has multiple indegree/outdegree. To correctly find cognates, not only the meaning (which is predicted by shared edges) but also the form need to be considered. We add form-similarity/lexical distance rate as a new heuristic in finding cognates following [17] using the Longest Common Subsequence Ratio (LCSR).

Some linguistic studies show that the meaning of a word can be deduced via cognate synonym [9, 34]. For instance, in Figure 3, w_1^A, w_2^A, and w_3^A are words in the Minangkabau language (min); w_1^B, w_2^B and w_3^B are words in the Indonesian language (ind); and w_1^C, w_2^C, and w_3^C are words in the Malay language (zlm). When we connect words in non-pivot language A and C via pivot words B based on shared meaning between the words, we can get translation results from language A to C. In this example, we have information about senses/meanings for all words in input dictionaries and there are three cognates that are (w_1^A, w_1^B, w_1^C), (w_2^A, w_2^B, w_2^C), and (w_3^A, w_3^B, w_3^C), as indicated



For example, from step 1, cognate pair is identified: $w_1^A - w_2^C$
Step 2 – Recognize cognate synonyms:
a. Recognize synonyms of w_2^C based on ratio of shared connectivity with the pivot word(s):
- Probability of $w_2^C - w_3^C$ being synonym: 3/3 = 1
- Probability of $w_2^C - w_4^C$ being synonym: 2/3 = 0.67
- Probability of $w_2^C - w_1^C$ being synonym: 1/3 = 0.33

b. Pair w_1^A with the synonyms of w_2^C as cognate synonym pairs:
- $w_1^A - w_3^C, w_1^A - w_4^C, w_1^A - w_1^C$

Fig. 4. Cognate synonym recognition.

within the same box in Figure 3. A cognate w_1^A-w_1^C and non-cognates w_1^A-w_2^C and w_1^A-w_3^C are correct translations, since w_1^C, w_2^C, and w_3^C are synonymous.

Nevertheless, it remains a challenge to find the cognate synonyms when the input dictionaries do not have information about senses/meanings. As shown in Figure 4, to recognize cognate synonyms, first, we need to recognize synonyms of w_2^C based on ratio of shared connectivity with the pivot word(s), since we assume that synonymous words are connected to common pivot word(s). Then w_1^A will be paired with the recognized synonyms of w_2^C to obtain cognate synonym pairs. The higher the ratio of shared connectivity between a synonym of w_2^C with the pivot words (w_1^B, w_2^B, w_3^B), the higher the probability of the synonym being a translation pair with w_1^A.

Finally, by recognizing both cognate pairs and cognate synonym pairs, we can obtain many-to-many translation results.

5 GENERALIZATION OF CONSTRAINT-BASED LEXICON INDUCTION FRAMEWORK

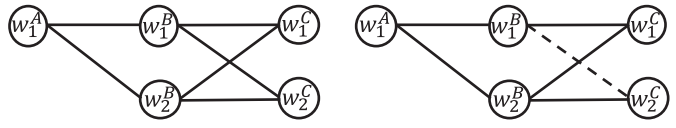
We generalize the constraint-based lexicon induction framework by extending the existing one-cycle symmetry assumption into the n-cycle symmetry assumption and identify cognates and cognate synonyms by utilizing four heuristics to improve the quality and quantity of the translation pair results.

5.1 Tripartite Transgraph

To model translation connectivity between language A and C via pivot language B, we define the tripartite transgraph, which is a tripartite graph in which a vertex represents a word and an edge represents the indication of shared meaning(s) between two vertices. Two tripartite transgraphs can be joined if there exists at least one edge connecting a pivot vertex in one tripartite transgraph to one non-pivot vertex in the other tripartite transgraph. To maintain the basic form of a tripartite transgraph with n number of pivot words (at least 1 pivot per transgraph), each pivot word must be connected to at least one word in every non-pivot language, and there has to be a path connecting all pivot words via non-pivot words. Hereafter, we abbreviate the tripartite transgraph to transgraph.

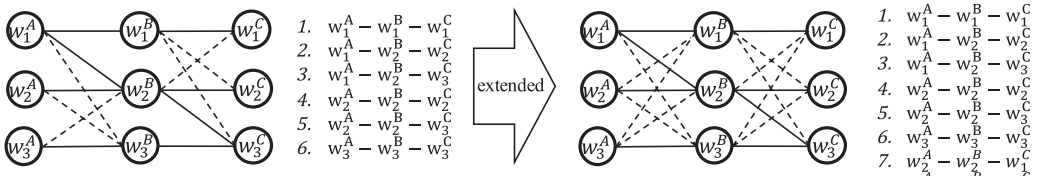
In this research, we assume that the input dictionaries contain no sense information. After modeling the translation connectivity from the input dictionaries as transgraphs, we further analyze the shared edges between the non-pivot vertices and the pivot vertices to predict the shared meanings between them. We then formalize the problem into Conjunctive Normal Form (CNF) and using WPMaxSAT solver to return the most probable correct translation results.

Sometimes, for high-resource languages where the input dictionaries have many shared meanings via the pivot words, a big transgraph can be generated, which potentially leads to excessive



(a) Symmetry Transgraph (b) Asymmetry Transgraph

Fig. 5. Symmetry and asymmetry transgraphs.



(a) Translation pair candidates from **existing solid-edges** in 1-cycle symmetry assumption

(b) Translation pair candidates from **existing solid edges** and ***previously added dashed-edges*** in 2-cycle symmetry assumption

Fig. 6. N-cycle symmetry assumption extension.

computational complexity when we formalize and solve it. Nevertheless, for low-resource languages where we can expect the input dictionaries to have just a few shared meanings via the pivot words, transgraph size is small enough to make its formalization and solution feasible. Therefore, for the sake of simplicity, we ignore big transgraphs in our experiments.

5.2 N-cycle Symmetry Assumption

Machine-readable bilingual dictionaries are rarely available for low-resource languages like Indonesian ethnic languages. It is even difficult to find sizable printed bilingual dictionary with acceptable quality for Indonesian ethnic languages. In the currently available machine-readable or printed dictionaries, we can expect to find missed senses/meanings that would lead to asymmetry in the transgraph. The expected missed senses are represented as dashed edges in the transgraph as depicted in Figure 5(b). The one-to-one approach only considers translation pair candidates from existing connected solid edges in the transgraph as shown in Figure 6(a). To fully satisfy symmetry constraint in the transgraph, we extend the existing one-cycle symmetry assumption to the n-cycle symmetry assumption while considering new translation pair candidates from the new dashed edges. As shown in Figure 6(b), during the second cycle, the previously new dashed edges developed in the first cycle are taken to exist, therefore, we can extract translation pair candidates not only from the solid edges but also from the previously added dashed-edges. Users can input the maximum cycle for the experiment as shown in Algorithm 2 (as *maxCycle*).

5.3 Formalization

Constraint optimization problem formalism has been used in solving many natural language processing and web service composition related problems [10, 16]. Our team [37] formalized bilingual lexicon induction as a WPMaxSAT problem. In this article, we follow the same formulation. A literal is either a Boolean variable x or its negation $\neg x$. A clause C is a disjunction of literals $x_1 \vee \ldots \vee x_n$. A unit clause is a clause consisting of a single literal. A weighted clause is a pair (C, ω), where C is a clause and ω is a natural number representing the penalty for falsifying the clause C. If a clause is hard, then the corresponding weight is infinity. The propositional formula φ_c^ω in CNF [2] is a conjunction of one or more clauses $C_1 \wedge \ldots \wedge C_n$. CNF formula with soft clauses

is represented as φ_c^+ and φ_c^∞ represents a CNF formula with hard clauses. The WPMaxSAT problem for a multiset of weighted clauses C is the problem of finding an optimal assignment to the variables of C that minimizes the cost of the assignment on C. Let w_i^A, w_j^B, and w_k^C represents words from language A, B, and C. We define seven propositions as Boolean variables between a pair of words w_i^A, w_j^B, and w_k^C as follows:

- $e(w_i^A, w_j^B)$ and $e(w_j^B, w_k^C)$ represents edge existence between word pair from language A and B and from language B and C, respectively,
- $c(w_i^A, w_k^C)$, $c(w_i^A, w_n^C)$, and $c(w_m^A, w_k^C)$ represents whether the word pair from language A and C is a cognate pair, and
- $s(w_i^A, w_n^C)$ and $s(w_m^A, w_k^C)$ represents whether the word pair from language A and C is a cognate synonym pair

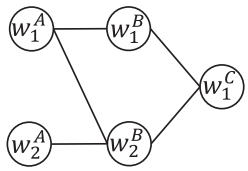
To encode some of the constraints to CNF, we use a resolution approach based on the Boolean algebra rule of $p \rightarrow q \wedge r \Leftrightarrow (\neg p \vee q) \wedge (\neg p \vee r)$. In the framework, we define E_E as a set of word pairs connected by existing edges, E_N as a set of word pairs connected by new edges, D_C as a set of translation pair candidates, D_{Co} as a set of cognate pairs, D_{NCo} as a set of non-cognate pairs, D_{PCo} as a set of pivot words from language B that are connecting the current cognate pair, and D_R as a set of all translation pair results returned by the WPMaxSAT solver.

5.4 Heuristics to Find Cognate

We define four heuristics to find cognates in the transgraph: cognate pair coexistence probability, missing contribution rate toward cognate pair coexistence, polysemy pivot ambiguity rate, and cognate form similarity. Based on our symmetry assumption, when w_i^A and w_k^C in a transgraph share all of their senses through pivot word(s) from language B, they are a potential cognate pair, where the cognate pair coexistence probability equals 1, the missing contribution equals 0 and the polysemy pivot ambiguity rate equals 0. When w_i^A and w_k^C have the same spelling, they are a potential cognate pair, where the cognate form similarity equals 1. Thus, when w_i^A and w_k^C are satisfying the symmetry assumption and also have the same spelling, we take them as the highest potential cognate pair in the transgraph.

5.4.1 Cognate Pair Coexistence Probability. Cognate pairs of language A and C are induced from two input bilingual dictionaries, that is, Dictionary A-B and Dictionary B-C. We define two sets of event for Dictionary A-B (w_i^A and w_j^B) where event w_i^A represents connecting word w_i^A of language A to words of language B represented by edges based on shared meaning between them. Similarly, event w_j^B represents connecting word w_j^B of language B to words of language A. We also define two sets of event for Dictionary B-C (w_j^B and w_k^C), where event w_j^B represents connecting word w_j^B of language B to words of language C and event w_k^C represents connecting word w_k^C of language C to words of language B. A marginal probability $P(w_i^A)$ is a probability of w_i^A connected to words of language B. A conditional probability $P(w_i^A|w_j^B)$ is a probability of w_i^A connected to w_j^B considering other words of language A that also connected to w_j^B. A joint probability $P(w_i^A, w_j^B)$ is a probability of w_i^A interconnected to w_j^B. For example, in Figure 7, $P(w_1^A) = 2/3$, since w_1^A has two connected edges with words of language B of three existing connected edges between words of language A and words of language B. The joint probability $P(w_1^A, w_1^B) = 1/3$, since any word from language A and any word from language B are only interconnected with one edge of three existing connected edges between words of language A and words of language B.

To calculate the possibility of a translation pair candidate $t(w_i^A, w_k^C)$ being a cognate pair $c(w_i^A, w_k^C)$, we calculate $t(w_i^A, w_k^C).H_{coex}$, a cognate coexistence probability of translation pair

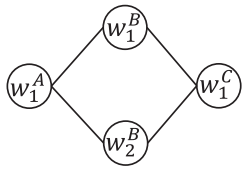


	w_1^B	w_2^B	Total
w_1^A	$1/3^*$	$1/3^*$	$2/3^{**}$
w_2^A	0^*	$1/3^*$	$1/3^{**}$
Total	$1/3^{**}$	$2/3^{**}$	1

	w_1^C	Total
w_1^B	$1/2^*$	$1/2^{**}$
w_2^B	$1/2^*$	$1/2^{**}$
Total	1^{**}	1

* Joint Probability $P(w_i^X, w_j^Y)$
** Marginal Probability $P(w_i^X)$

Fig. 7. Example of marginal and joint probability.



	w_1^B	w_2^B	Total
w_1^A	1/2	1/2	1
Total	1/2	1/2	1

	w_1^C	Total
w_1^B	1/2	1/2
w_2^B	1/2	1/2
Total	1	1

$P(w_1^A, w_1^C) = P(w_1^A|w_1^C)P(w_1^C|w_1^A)$
$= (P(w_1^A|w_1^B)P(w_1^B|w_1^C) + P(w_1^A|w_2^B)P(w_2^B|w_1^C))\ (P(w_1^C|w_1^B)P(w_1^B|w_1^A) + P(w_1^C|w_2^B)P(w_2^B|w_1^A))$
$= (1/2 * 1/2 + 1/2 * 1/2)(1 * 1/2 + 1 * 1/2)$
$= 1$

Fig. 8. Symmetry pair coexistence probability.

candidate $t(w_i^A, w_k^C)$. We first utilize a chain rule to obtain Equations (1) and (2). By multiplying them, we can get Equation (3). Event w_i^A and event w_k^C are independent, since they are from a different input bilingual dictionary, thus, $P(w_k^C, w_i^A) = P(w_i^A)P(w_k^C)$ and Equation (3) can be rewritten as Equation (4). We use a generative probabilistic process commonly used in prior work [6, 18, 22, 35] in Equation (5) to obtain $P(w_i^A|w_k^C)$ and $P(w_k^C|w_i^A)$. Finally, we can obtain a cognate coexistence probability of translation pair candidate $t(w_i^A, w_k^C)$ as $t(w_i^A, w_k^C).H_{coex} = P(w_i^A, w_k^C)$.

$$P(w_i^A, w_k^C) = P(w_k^C|w_i^A)P(w_i^A), \tag{1}$$

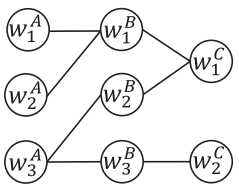
$$P(w_k^C, w_i^A) = P(w_i^A|w_k^C)P(w_k^C), \tag{2}$$

$$P(w_i^A, w_k^C)P(w_k^C, w_i^A) = P(w_i^A|w_k^C)P(w_k^C|w_i^A)P(w_i^A)P(w_k^C), \tag{3}$$

$$P(w_i^A, w_k^C) = P(w_i^A|w_k^C)P(w_k^C|w_i^A), \tag{4}$$

$$P(w_i^A|w_k^C) = \sum_{j=0} P(w_i^A|w_j^B)P(w_j^B|w_k^C). \tag{5}$$

When w_i^A and w_k^C in a transgraph share all of their senses through pivot word(s) from language B and none of the pivot words are ambiguous, the cognate pair coexistence probability equals 1, as shown in Figure 8. The algorithm to calculate the probability of the translation pair candidates coexisting as a cognate is shown in Algorithm 1 line number 19. The coexistence probability is very important in differentiating cognates from non-cognates, but it is poor at avoiding polysemy in pivot words. This is because it treats polysemy in the pivot words and polysemy in the non-pivot words equally. In reality, however, polysemy in pivot words negatively impacts the quality



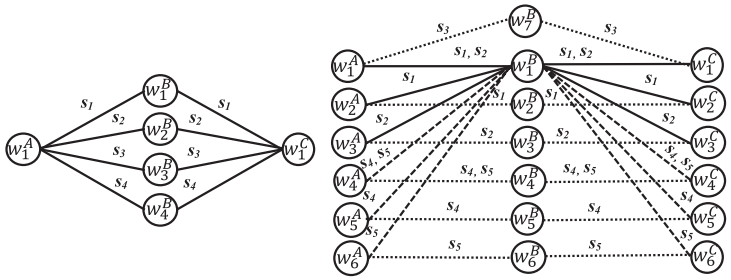
	w_1^B	w_2^B	w_3^B	Total
w_1^A	1/4	0	0	1/4
w_2^A	1/4	0	0	1/4
w_3^A	0	1/4	1/4	2/4
Total	2/4	1/4	1/4	1

	w_1^C	w_2^C	Total
w_1^B	1/3	0	1/3
w_2^B	1/3	0	1/3
w_3^B	0	1/3	1/3
Total	2/3	1/3	1

$P(w_1^A, w_1^C) = P(w_1^A|w_1^C)P(w_1^C|w_1^A)$
$= (P(w_1^A|w_1^B)P(w_1^B|w_1^C))\ (P(w_1^C|w_1^B)P(w_1^B|w_1^A))$
$= (1/2 * 1/2)(1 * 1)$
$= 1/4$

$P(w_3^A, w_1^C) = P(w_3^A|w_1^C)P(w_1^C|w_3^A)$
$= (P(w_3^A|w_1^B)P(w_1^B|w_1^C))\ (P(w_1^C|w_1^B)P(w_1^B|w_3^A))$
$= (1 * 1/2)(1 * 1/2)$
$= 1/4$

Fig. 9. Equal treatment of polysemy in pivot/non-pivot word.



(a) Polysemy in non-pivot language (b) Polysemy in pivot language

Fig. 10. Polysemy in pivot and non-pivot language.

of bilingual dictionary induction rather than polysemy in non-pivot words. A case with high polysemy in pivot words and low polysemy in non-pivot words and a case with low polysemy in pivot words and high polysemy in non-pivot words where the two cases have equal rates of polysemy, will yield same probability as shown in Figure 9. Therefore, we introduce a special heuristic to tackle this problem, that is, polysemy pivot ambiguity rate.

5.4.2 Missing Contribution Rate Toward Cognate Pair Coexistence. Inspired by the Shapley Value [26], a solution concept in cooperative game theory, we calculate missing contribution rate toward cognate pair coexistence probability by calculating coexistence probability of supposed cognate pair (also considering missing edges as existing) minus the coexistence probability of the pair from existing connectivity only. When w_i^A and w_k^C in a transgraph share all of their senses through pivot word(s) from language B (no missing senses), the missing contribution equals 0. The lower is the missing contribution toward coexistence probability of a translation pair candidate, the more likely is the translation pair candidate of being a cognate. The calculation of missing contribution rate of w_1^A and w_1^C pair, that is, $t(w_i^A, w_k^C).H_{missCont}$ is shown in Algorithm 1 line number 20.

5.4.3 Polysemy Pivot Ambiguity Rate. To model the effect of polysemy in the pivot language on precision, for the sake of simplicity, we ignore synonym words within the same language. Polysemy in non-pivot languages have no negative effect on precision. In Figure 10(a), even though the non-pivot words are connected by four pivot words representing four senses/meanings, the transgraph only has one translation pair candidate (w_1^A-w_1^C) and so the precision is 100%.

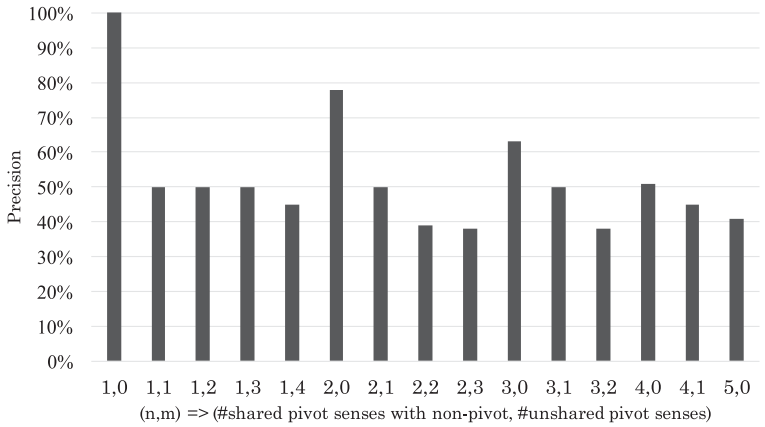


Fig. 11. Prediction model of precision on polysemy in pivot language.

However, polysemy in pivot language negatively impacts precision. Figure 10(b) shows that non-pivot word w_1^A and w_1^C are cognates and share the same meanings (s_1, s_2, s_3), but pivot word w_1^B that has four meanings (s_1, s_2, s_4, s_5) only shares a part of the meanings (s_1, s_2) with the non-pivot words. The solid edges have part or all shared meanings (s_1, s_2) between the non-pivot words (w_1^A, w_1^C) and the pivot word w_1^B. The dashed edges express part or all unshared meanings (s_4, s_5) between the non-pivot words (w_1^A, w_1^C) and the pivot word w_1^B. To investigate the effect of pivot word w_1^B on the overall precision, we extract only translation pair candidates from the connected edges. The precision (38.89%) is affected negatively as there are 22 wrong translations because of the polysemy in pivot language (w_1^B) in the transgraph.

We formalize the effect of polysemy in pivot language on precision with the following formulation where n is the number of shared meanings between pivot word and non-pivot words and m is the number of pivot meaning(s) that are not shared with non-pivot words. The number of correct translations contributed by the solid edges and the number of correct translations contributed by the dashed edges can be calculated by Equation (6). The precision of the translation result is calculated by Equation (7),

$$CorrectTrans(n) = 2\sum_{i=1}^{n}\sum_{j=1}^{i}\binom{n}{i}\binom{i}{j} - \sum_{i=1}^{n}\binom{n}{i}, \tag{6}$$

$$Precision(n,m) = \frac{CorrectTrans(n) + CorrectTrans(m)}{\left[\sum_{i=1}^{n}\binom{n}{i} + \sum_{i=1}^{m}\binom{m}{i}\right]^2}. \tag{7}$$

We predict the effect of shared meanings between pivot word and non-pivot words by simulating 10 sets of transgraphs with n (the number of shared meanings between pivot word and non-pivot words) values ranging from 1 to 10 where, in each set, m (the number of pivot meaning(s) that not shared with non-pivot words) ranges from 0 to n in Figure 11. In this experiment, non-pivot languages and pivot language are closely related languages (w_1^A, w_1^B, and w_1^C are cognates) when there is no pivot meaning that not shared with non-pivot words ($m = 0$). This result shows that the greater the number of shared senses/meanings (represented by n) between pivot and non-pivot words there are, the lower the precision is. Nevertheless, the polysemy in the pivot language has the least negative effect on the precision when the pivot language and non-pivot languages

are closely related where the number of unshared pivot senses (represented by m) equals 0. The negative effect increases as the value of m increases.

Polysemy in pivot words negatively impacts the precision of the translation result, unlike that in non-pivot words. Since we do not have any information about the senses from the input dictionaries, it is difficult to avoid the negative effect of the polysemous pivot word. To predict a probability of w_i^A and w_k^C to be a cognate pair via the pivot word w_j^B that shares common senses, we assume the worst-case scenario where the number of senses belonging to pivot word w_j^B equals the maximum number of connected edges to w_i^A or w_k^C. If the maximum number of indegree or outdegree of the polysemy pivot is n, then there are $2^n - 1$ possible combination of shared senses for every paths via pivot word w_j^B in order for the translation pair candidates to be a cognate pair $c(w_i^A, w_k^C)$ of all $(2^n - 1)^2$ combinations. In Figure 5(b), the possible combination of shared senses between w_1^A and w_1^C or between w_1^A and w_2^C are $[s_1, s_2, s_1 \,\&\, s_2]$. To calculate the probability of the pair w_i^A and w_k^C being a cognate considering polysemy in the pivot words, we calculate $t(w_i^A, w_k^C).P_{sharedSenses}$, the product of the probabilities of shared senses between the pair for every existing path as shown in Algorithm 1 line number 10. The polysemy pivot ambiguity rate is given by $t(w_i^A, w_k^C).H_{polysemy} = 1 - t(w_i^A, w_k^C).P_{sharedSenses}$ as shown in Equation (8) and Algorithm 1 line number 21,

$$t(w_i^A, w_k^C).H_{polysemy} = 1 - \prod \left((2^n - 1)/(2^n - 1)^2\right) = 1 - \prod \left(1/(2^n - 1)\right). \tag{8}$$

The lower the polysemy pivot ambiguity rate is, the more likely it is that the pair form a cognate and share exact senses. When there is only one path between w_i^A and w_k^C and there is only one indegree and one outdegree of the pivot word w_j^B, the polysemy pivot ambiguity rate equals 0.

5.4.4 Cognate Form Similarity. Because the symmetry assumption can sometimes fail to select a cognate correctly when it gives the same cost for multiple translation pair candidates, the cognate form similarity heuristic will contribute to selecting the cognate. We calculate cognate form similarity using LCSR ranging from 0 (0% form-similarity) to 1 (100% form-similarity) following Reference [17] as shown in Equation (9) and Algorithm 1 line number 22, where $LCS(w_i^A, w_k^C)$ is the longest common subsequence of w_i^A and w_k^C; $|x|$ is the length of x; and $max(|w_i^A|, |w_k^C|)$ returns the longest length. However, the maximum cost contributed from the form dissimilarity is set at 1/100 of the maximum cost contributed by one symmetry assumption heuristic as shown in Algorithm 2 line number 24 to ensure that the cognate form similarity heuristic will have only a supporting role in helping the main symmetry assumption heuristics,

$$LCSR(w_i^A, w_k^C) = \frac{|LCS(w_i^A, w_k^C)|}{max(|w_i^A|, |w_k^C|)}, \tag{9}$$

$$t(w_i^A, w_k^C).H_{formSim} = LCSR(w_i^A, w_k^C). \tag{10}$$

5.5 Constraints Extension

We extend the one-to-one approach constraints by adding several new constraints to the constraint sets to find cognates and cognate synonyms. All constraints are listed in Table 2.

5.5.1 Edge Existence. An edge exists in the transgraph between words that share their meaning(s) based on input dictionaries. The existing edges in the transgraph are encoded as TRUE, that is, $e(w_i^A, w_j^B)$ and $e(w_j^B, w_k^C)$ in the CNF formula, which is represented as hard constraint φ_1^∞.

ALGORITHM 1: Cognate Pair Probability Calculation

Input: Translation pair candidate $t(w_i^A, w_k^C)$;
Output: Translation pair candidate $t(w_i^A, w_k^C)$ with cognate pair probabilites information

$P(w_i^A, w_k^C) = 0$; $P(w_k^C, w_i^A) = 0$; $P_{missing}(w_i^A, w_k^C) = 0$; $P_{missing}(w_k^C, w_i^A) = 0$;
for *each path in* $t(w_i^A, w_k^C).Paths$ **do**
  $P(w_i^A|w_j^B) = 0$; $P(w_j^B|w_k^C) = 0$; $P(w_k^C|w_j^B) = 0$; $P(w_j^B|w_i^A) = 0$;
  `/* Conditional Probability direction: A-C */`
  for *each inEdge in* $w_j^B.inEdges$ *from language A* **do** $w_j^B.indegreeFromA$ += $1/inEdge.Prob$;
  for *each inEdge in* $w_k^C.inEdges$ *from language B* **do** $w_k^C.indegreeFromB$ += $1/inEdge.Prob$;
  $P(w_i^A|w_j^B) = 1 / w_j^B.indegreeFromA$; $P(w_j^B|w_k^C) = 1 / w_k^C.indegreeFromB$;
  `/* Conditional Probability direction: C-A */`
  for *each inEdge in* $w_j^B.inEdges$ *from language C* **do** $w_j^B.indegreeFromC$ += $1/inEdge.Prob$;
  for *each inEdge in* $w_i^A.inEdges$ *from language B* **do** $w_i^A.indegreeFromB$ += $1/inEdge.Prob$;
  $P(w_k^C|w_j^B) = 1 / w_j^B.indegreeFromC$; $P(w_j^B|w_i^A) = 1 / w_i^A.indegreeFromB$;
  $t(w_i^A, w_k^C).P_{sharedSenses}$ *= $1/(2^{max(w_j^B.indegreeFromA, w_j^B.indegreeFromC)} - 1)$;
  if *missing edge exist in path* **then**
    $P_{missing}(w_i^A|w_k^C)$ += $P(w_i^A|w_j^B)P(w_j^B|w_k^C)$;
    $P_{missing}(w_k^C|w_i^A)$ += $P(w_k^C|w_j^B)P(w_j^B|w_i^A)$;
  else
    $P(w_i^A|w_k^C)$ += $P(w_i^A|w_j^B)P(w_j^B|w_k^C)$;
    $P(w_k^C|w_i^A)$ += $P(w_k^C|w_j^B)P(w_j^B|w_i^A)$;
  end
end
$t(w_i^A, w_k^C).H_{coex} = P(w_i^A|w_k^C)P(w_k^C|w_i^A)$;
$t(w_i^A, w_k^C).H_{missCont} = (P(w_i^A|w_k^C) + P_{missing}(w_i^A|w_k^C))(P(w_k^C|w_i^A) + P_{missing}(w_k^C|w_i^A)) - (P(w_i^A|w_k^C)P(w_k^C|w_i^A))$;
$t(w_i^A, w_k^C).H_{polysemy} = 1 - t(w_i^A, w_k^C).P_{sharedSenses}$;
$t(w_i^A, w_k^C).H_{formSim} = LCSR(w_i^A, w_k^C)$;
return $t(w_i^A, w_k^C)$;

5.5.2 Edge Non-Existence. An edge does not exist in the transgraph between words that do not share their meaning(s) based on input dictionaries. We formalize the non-existence of edge in the transgraph by encoding the negation of the literal edge existence, that is, $\neg e(w_i^A, w_j^B)$ and $\neg e(w_j^B, w_k^C)$ in the CNF formula, which is represented as soft constraint φ_2^+.

5.5.3 Symmetry. Cognate share all of their senses/meanings and symmetrically connected via pivot language B. We convert $c(w_i^A, w_k^C) \rightarrow e(w_i^A, w_1^B) \wedge e(w_i^A, w_2^B) \wedge \ldots \wedge e(w_1^B, w_k^C) \wedge e(w_2^B, w_k^C) \wedge \ldots$ into $(\neg c(w_i^A, w_k^C) \vee e(w_i^A, w_1^B)) \wedge (\neg c(w_i^A, w_k^C) \vee e(w_i^A, w_2^B)) \wedge \ldots. \wedge (\neg c(w_i^A, w_k^C) \vee e(w_1^B, w_k^C)) \wedge (\neg c(w_i^A, w_k^C) \vee e(w_1^B, w_k^C)) \wedge \ldots$ It is encoded as hard constraint φ_3^∞. Unfortunately, a problem arises with low-resource languages where the input dictionaries have no sense information and many senses are expected to be missed due to the small size of the dictionaries. To solve this problem, we add new edges so cognate pairs share all of the meanings by violating the edge non-existence soft constraint φ_2^+ and paying a cost determined from user-selected heuristics (cognate pair coexistence probability, missing contribution rate toward the cognate pair

Table 2. Constraints for Cognates and Cognate Synonyms Extraction

ID	CNF Formula
	Edge Existence:
φ_1^{∞}	$\Big(\bigwedge_{(w_i^A, w_j^B)\in E_E} \big(e(w_i^A, w_j^B), \infty\big)\Big) \wedge \Big(\bigwedge_{(w_j^B, w_k^C)\in E_E} \big(e(w_j^B, w_k^C), \infty\big)\Big)$
	Edge Non-Existence:
φ_2^{+}	$\Big(\bigwedge_{(w_i^A, w_j^B)\in E_N} \big(\neg e(w_i^A, w_j^B), \omega(w_i^A, w_j^B)\big)\Big) \wedge \Big(\bigwedge_{(w_j^B, w_k^C)\in E_N} \big(\neg e(w_j^B, w_k^C), \omega(w_j^B, w_k^C)\big)\Big)$
	Symmetry:
φ_3^{∞}	$\Big(\bigwedge_{\substack{(w_i^A, w_j^B)\in E_E\cup E_N \\ (w_i^A, w_k^C)\in D_C}} \big((\neg c(w_i^A, w_k^C) \vee e(w_i^A, w_j^B)), \infty\big)\Big) \wedge \Big(\bigwedge_{\substack{(w_j^B, w_k^C)\in E_E\cup E_N \\ (w_i^A, w_k^C)\in D_C}} \big((\neg c(w_i^A, w_k^C) \vee e(w_j^B, w_k^C)), \infty\big)\Big)$
	Uniqueness:
φ_4^{∞}	$\Big(\bigwedge_{\substack{k\neq n \\ (w_i^A, w_k^C)\in D_C \\ (w_i^A, w_n^C)\in D_C}} \big((\neg c(w_i^A, w_k^C) \vee \neg c(w_i^A, w_n^C)), \infty\big)\Big) \wedge \Big(\bigwedge_{\substack{i\neq m \\ (w_i^A, w_k^C)\in D_C \\ (w_i^A, w_n^C)\in D_C}} \big((\neg c(w_i^A, w_k^C) \vee \neg c(w_m^A, w_k^C)), \infty\big)\Big)$
	Extracting at Least One Cognate:
φ_5^{∞}	$\Big(\Big(\bigvee_{(w_i^A, w_k^C)\notin D_R} c(w_i^A, w_k^C)\Big), \infty\Big)$
	Encoding Cognate:
φ_6^{∞}	$\bigwedge_{(w_i^A, w_k^C)\in D_{Co}} \big(c(w_i^A, w_k^C), \infty\big)$
	Encoding Non-Cognate:
φ_7^{∞}	$\bigwedge_{(w_i^A, w_k^C)\in D_{NCo}} \big(\neg c(w_i^A, w_k^C), \infty\big)$
	Cognate Synonym:
φ_8^{∞}	$\Big(\bigwedge_{\substack{k\neq n \\ (w_i^A, w_k^C)\in D_{Co} \\ (w_i^A, w_n^C)\notin D_R}} \big((\neg s(w_i^A, w_n^C) \vee c(w_i^A, w_k^C)), \infty\big) \wedge \Big(\bigwedge_{w_j^B\in D_{PCo}} \big((\neg s(w_i^A, w_n^C) \vee e(w_j^B, w_n^C)), \infty\big)\Big)\Big)$
	$\wedge\Big(\bigwedge_{\substack{i\neq m \\ (w_m^A, w_k^C)\in D_{Co} \\ (w_i^A, w_k^C)\notin D_R}} \big((\neg s(w_m^A, w_k^C) \vee c(w_i^A, w_k^C)), \infty\big) \wedge \Big(\bigwedge_{w_j^B\in D_{PCo}} \big((\neg s(w_m^A, w_k^C) \vee e(w_m^A, w_j^B)), \infty\big)\Big)\Big)$
	Extracting at Least One Cognate Synonym:
φ_9^{∞}	$\Big(\Big(\bigvee_{(w_i^A, w_k^C)\notin D_R} s(w_i^A, w_k^C)\Big), \infty\Big)$

coexistence probability, polysemy pivot ambiguity rate, and cognate form similarity). In other words, we assume the edges exist. The higher the cognate pair coexistence probability and the lower the missing contribution rate toward the cognate pair coexistence probability and the lower the polysemy pivot ambiguity rate and the higher the cognate form similarity, the more likely it is that the pair form a cognate, thus, the lower is the cost of adding any new edge to it, that is, the new edge weight. The new edges in the transgraph is encoded as FALSE (NOT exist), that is, $\neg e(w_i^A, w_j^B)$ or $\neg e(w_j^B, w_k^C)$ in the CNF formula and depicted as dashed edges in the transgraph. The weight of the new edge from non-pivot word w_i^A to pivot word w_j^B is defined as $\omega(w_i^A, w_j^B)$ and the weight of a new edge from pivot word w_j^B to non-pivot word w_k^C is defined as $\omega(w_j^B, w_k^C)$.

Both of $\omega(w_i^A, w_j^B)$ and $\omega(w_j^B, w_k^C)$ values equal $t(w_i^A, w_k^C).H_{coex} + t(w_i^A, w_k^C).H_{missCont} + t(w_i^A, w_k^C).H_{polysemy} + t(w_i^A, w_k^C).H_{formSim}$ as shown in Algorithm 2 line numbers 21–24.

5.5.4 Uniqueness. The first step of our strategy in obtaining many-to-many translation pair with good quality is to extract a list of cognates in the transgraph. The uniqueness constraint ensures that only one-to-one cognates that share all of their meanings will be considered as translation pairs. In other words, a word in language A can only be a cognate with just one word from language C. This is encoded as hard constraint φ_4^∞.

5.5.5 Extracting at Least One Cognate. Since the framework communicates with WPMaxSAT solver iteratively as shown in Algorithm 2 line numbers 2–7, hard constraint φ_5^∞ ensures that at least one of the $c(w_i^A, w_k^C)$ variables must be evaluated as TRUE. Consequently, each iteration yields one most probable cognate pair and stores it in D_{Co} and also in D_R as a translation pair result. This clause is a disjunction of all $c(w_i^A, w_k^C)$ variables.

5.5.6 Encoding Cognate. We exclude previously selected translation pairs, which are stored in D_{Co} from the following list of cognate pair candidates by encoding them as TRUE, that is, $c(w_i^A, w_k^C)$, which is encoded as hard constraint φ_6^∞, and excluding them from φ_5^∞.

5.5.7 Encoding Non-Cognate. Once we get list of all cognate pairs, stored in D_{Co}, the remaining translation pair candidates are stored in D_{NCo} and encoded as FALSE, that is, $\neg c(w_i^A, w_k^C)$ in the CNF formula, which is represented as hard constraint φ_7^∞.

5.5.8 Cognate Synonym. We can further identify cognate synonyms to improve the quantity of the translation results. For each cognate pair $c(w_i^A, w_k^C)$ stored in D_{Co}, we can find cognate synonym pairs $s(w_i^A, w_n^C)$ and $s(w_m^A, w_k^C)$ by extracting synonyms of w_k^C and w_i^A, respectively. We assume that synonymous words are connected to common pivot words. We can add new edges by paying cost of violating soft-constraint φ_2^+ with a weight different from that used when identifying cognate pairs in the first step. In this second step, the weight is calculated based on cognate synonym probability $P_{cognateSyn}$ for both w_n^C-w_k^C and w_m^A-w_i^A based on percentage of shared connectivity with the pivot words. The weight, that is, $1 - P_{cognateSyn}$ is distributed evenly to each new edges. We convert $s(w_i^A, w_n^C) \rightarrow c(w_i^A, w_k^C) \wedge e(w_1^B, w_n^C) \wedge e(w_2^B, w_n^C) \wedge \ldots$ into $(\neg s(w_i^A, w_n^C) \vee c(w_i^A, w_k^C)) \wedge (\neg s(w_i^A, w_n^C) \vee e(w_1^B, w_n^C)) \wedge (\neg s(w_i^A, w_n^C) \vee e(w_2^B, w_n^C)) \wedge \ldots$. With the same rule, we convert $s(w_m^A, w_k^C) \rightarrow c(w_i^A, w_k^C) \wedge e(w_m^A, w_1^B) \wedge e(w_m^A, w_2^B) \wedge \ldots$ into $(\neg s(w_m^A, w_k^C) \vee c(w_i^A, w_k^C)) \wedge (\neg s(w_m^A, w_k^C) \vee e(w_m^A, w_1^B)) \wedge (\neg s(w_m^A, w_k^C) \vee e(w_m^A, w_2^B)) \wedge \ldots$. It is encoded as hard constraint φ_8^∞. In Figure 4, $s(w_1^A, w_3^C).P_{cognateSyn} = 1$, $s(w_1^A, w_4^C).P_{cognateSyn} = 0.67$, and $s(w_1^A, w_1^C).P_{cognateSyn} = 0.33$. Another example, in Figure 5(a), if cognate pair $c(w_1^A, w_1^C)$ is identified, we need to identify cognate synonym probability of w_1^A (no candidate exist) and w_1^C (candidate: w_2^C). Based on the rate of shared connectivity with pivot word(s), $s(w_1^A, w_2^C).P_{cognateSyn} = 2/2$ and in Figure 5(b) with the same way we can get $s(w_1^A, w_2^C).P_{cognateSyn} = 1/2$.

5.5.9 Extracting at Least One Cognate Synonym. In the second step, that is, finding cognate synonyms, the framework also communicates with the WPMaxSAT solver iteratively as shown in Algorithm 2 line numbers 8–13, and hard constraint φ_9^∞ ensures that at least one of the $s(w_i^A, w_n^C)$ variables or $s(w_m^A, w_k^C)$ variables must be evaluated as TRUE. Consequently, each iteration yields one most probable cognate synonym pair and store it in D_R as a translation pair result. This clause is a disjunction of all $s(w_i^A, w_k^C)$ variables.

ALGORITHM 2: Cognate and Cognate Synonym Extraction

Input: *transgraphs*, *maxCycle*, *threshold*, *HSelections*;
Output: D_R /* list of translation pair results */
for *each transgraph in* calculateEdgeCost(*transgraphs*)) **do**
  /* Extract the most probable cognate pair and cognate synonym pair with total cost of violating constraints below the threshold iteratively */
  $CNF_{cognate} \leftarrow constructCNF_{cognate}(transgraph.D_C)$; /* following Equation (11) */
  while $cognatePair \leftarrow SATSolver.solve(CNF_{cognate})$ **do**
    if $cognatePair.totalCost < cognateThreshold$ **then**
      $D_R \leftarrow cognatePair$; $CNF_{cognate}.update()$;
    end
  end
  $CNF_{cognateSynonym} \leftarrow constructCNF_{cognateSynonym}(transgraph.D_C)$; /* following Equation (12) */
  while $cognateSynonymPair \leftarrow SATSolver.solve(CNF_{cognateSynonym})$ **do**
    if $cognateSynonymPair.totalCost < cognateSynonymThreshold$ **then**
      $D_R \leftarrow cognateSynonymPair$; $CNF_{cognateSynonym}.update()$;
    end
  end
end
return D_R;
Function calculateEdgeCost(*transgraphs*)
  for *each transgraph in transgraphs* **do**
    $transgraph.D_C \leftarrow generateCandidates(transgraph)$; /* generate trans. pair cand. */
    for *each* $t(w_i^A, w_k^C)$ *in* $transgraph.D_C$ **do**
      calculateCognatePairProb($t(w_i^A, w_k^C)$); /* using Algorithm 1 */
      /* Cost of adding new edges are calculated from user selected heuristics */
      if *HSelections.coex is TRUE* **then** $t(w_i^A, w_k^C).EdgeCost \mathrel{+}= 1 - t(w_i^A, w_k^C).H_{coex}$;
      if *HSelections.missCont is TRUE* **then** $t(w_i^A, w_k^C).EdgeCost \mathrel{+}= t(w_i^A, w_k^C).H_{missCont}$;
      if *HSelections.polysemy is TRUE* **then** $t(w_i^A, w_k^C).EdgeCost \mathrel{+}= t(w_i^A, w_k^C).H_{polysemy}$;
      if *HSelections.formSim is TRUE* **then**
      $t(w_i^A, w_k^C).EdgeCost \mathrel{+}= \left(1 - t(w_i^A, w_k^C).H_{formSim}\right)/100$;
      for *each* $w_j^A.outEdges$ **do**
        if $e(w_j^B, w_k^C)$ *is not exist* **then** $t(w_i^A, w_k^C).e(w_j^B, w_k^C).Cost = t(w_i^A, w_k^C).EdgeCost$;
      end
      for *each* $w_i^C.inEdges$ **do**
        if $e(w_i^A, w_j^B)$ *is exist* **then** $t(w_i^A, w_k^C).e(w_i^A, w_j^B).Cost = t(w_i^A, w_k^C).EdgeCost$;
      end
    end
  end
  if *maxCycle is not reached* **then**
    $transgraphs \leftarrow addNewEdges()$; /* add new edges to transgraphs for the next cycle */
    calculateEdgeCost(*transgraphs*);
  end
  return *transgraphs*

Table 3. Variation of Constraint-based Bilingual Dictionary Induction

Cycle	$CNF_{cognate}$	$CNF_{cognate} + CNF_{cognateSynonym}$	CNF_{M-M}
1	H1[1], H2, H3, H4, H12, ...	H1, H2, H3, H4, H12, ...	H1[2]
>1	H1, H2, H3, H4, H12, ...	H1, H2, H3, H4, H12, ...	H1[3]

[1] *Identical to one-to-one approach [36] and Ω_1 in our prior work [20].*
[2] *Identical to Ω_2 in our prior work [20].*
[3] *For 2-cycle, identical to Ω_3 in our prior work [20].*

5.6 Framework Generalization

We define two main CNF formulas; one for recognizing cognate pairs, that is, $CNF_{cognate}$ as shown in Equation (11) and one for recognizing cognate synonym pairs, that is, $CNF_{cognateSynonym}$ as shown in Equation (12). We also define another CNF formula, that is, CNF_{M-M} as shown in Equation (13), which extracts many-to-many translation pairs by ignoring uniqueness constraint of the one-to-one approach [20]. Three constraints are shared by the CNF formulas: φ_1^{∞}, φ_2^{+}, and φ_6^{∞},

$$CNF_{cognate} = \varphi_1^{\infty} \wedge \varphi_2^{+} \wedge \varphi_3^{\infty} \wedge \varphi_4^{\infty} \wedge \varphi_5^{\infty} \wedge \varphi_6^{\infty}, \tag{11}$$

$$CNF_{cognateSynonym} = \varphi_1^{\infty} \wedge \varphi_2^{+} \wedge \varphi_6^{\infty} \wedge \varphi_7^{\infty} \wedge \varphi_8^{\infty} \wedge \varphi_9^{\infty}, \tag{12}$$

$$CNF_{M-M} = \varphi_1^{\infty} \wedge \varphi_2^{+} \wedge \varphi_3^{\infty} \wedge \varphi_5^{\infty} \wedge \varphi_6^{\infty}. \tag{13}$$

Various constraint-based bilingual dictionary induction methods can be constructed to suit different situations and purposes by using a cognate recognition ($CNF_{cognate}$) or a cognate & cognate synonym recognition ($CNF_{cognate} + CNF_{cognateSynonym}$) methods with a choice of n-cycle symmetry assumption, and with a series of individual and combined heuristics to be chosen as shown in Table 3. We can also define many-to-many translation pair extraction method in our previous work using CNF_{M-M}. Thus, we define our methods using Backus Normal Form as follows:

⟨*situatedMethod*⟩ ::= ⟨*cycle*⟩" : "⟨*method*⟩" : "⟨*heuristic*⟩
⟨*cycle*⟩ ::= "1"|"2"|"3"|"4"|"5"|"6"|"7"|"8"|"9"
⟨*method*⟩ ::= "*C*"|"*S*"|"*M*"
⟨*heuristic*⟩ ::= "*H*1"|"*H*2"|"*H*3"|"*H*4"|"*H*12"|"*H*13"|"*H*14"|"*H*23"|"*H*24"|"*H*123"|"*H*124"|"*H*234"

- *cycle*: symmetry assumption cycle where cycle ≥ 1.
- *method*: *C* as a cognate recognition ($CNF_{cognate}$) or *S* as a cognate & cognate synonym recognition ($CNF_{cognate} + CNF_{cognateSynonym}$) or *M* as a many-to-many approach (Ω_2 & Ω_3) in our previous work [20].
- *heuristic*: an individual or combined heuristics where H1234 means a combination of heuristic 1 (cognate pair coexistence probability), heuristic 2 (missing contribution rate toward cognate pair coexistence), heuristic 3 (polysemy pivot ambiguity rate), and heuristic 4 (cognate form similarity).

A combination of cognate only ($CNF_{cognate}$) method with 1-cycle symmetry assumption and heuristic 1 is defined as 1:C:H1, yielding an identical method with one-to-one approach [36] and Ω_1 in our prior work [20]. A combination of cognate only (CNF_{M-M}) method with heuristic 1 and 1-cycle symmetry assumption is defined as 1:M:H1, which is identical with Ω_2 and for 2-cycle symmetry assumption is defined as 2:M:H1, which is identical with Ω_3 in our prior work [20].

Table 4. Language Similarity of Input Dictionaries

Language Pair	Language Similarity
min-ind, zlm-ind, min-zlm	69.14%, 87.70%, 61.66%
deu-eng, nld-eng, deu-nld	31.38%, 39.27% 51.17%
spa-eng, por-eng, spa-por	6.66%, 3.79%, 32.04%
deu-eng, ita-eng, deu-ita	31.38%, 9.75%, 13.64%

6 EXPERIMENT

To evaluate our result, we calculate precision, recall, and the harmonic mean of precision and recall using the traditional F-measure or balanced F-score [23]. In each iteration, WPMaxSAT solver returns the optimal translation pair result with minimum total cost (incurred by violating some soft constraints). Translation pair result with total cost above the threshold are not considered. For the methods equivalent with our prior work [20], which are 1:C:H1, 1:M:H1, and 2:M:H1, we do not set any threshold. We try to analyze the impact of the threshold and the heuristics on the precision, recall, and F-score. For this purpose, we need to have a Gold Standard, so for each experiment, we can iterate threshold from 0 to the highest cost of constraint violation cost with 0.01 interval and try every combination of heuristics as input to Algorithm 2 (as *threshold* & *HSelections*) while observing the resulting precision, recall or F-score after evaluation against the gold standard. In this article, we choose the result with the highest F-score. We want to analyze the algorithm so our generalized constraint approach can be applied to other datasets for various languages. We conduct experiments with six methods constructed from our generalized constraint approach in which three of them yielding one-to-one translation pairs (1-1), that is, Cognates recognition with all combination of heuristic and 1-cycle symmetry assumption (1:C:⟨*heuristic*⟩), 2-cycles symmetry assumption (2:C:⟨*heuristic*⟩), and 3-cycles symmetry assumption (3:C:⟨*heuristic*⟩), and the rest yielding many-to-many translation pairs (M-M), that is, Cognate and Cognate Synonyms recognition with all combination of heuristic and 1-cycle symmetry assumption (1:S:⟨*heuristic*⟩), 2-cycles symmetry assumption (2:S:⟨*heuristic*⟩), and 3-cycles symmetry assumption (3:S:⟨*heuristic*⟩). As baselines, we use three methods from our previous work where H1 is the sole heuristic used [20], that is, one-to-one translation pair extraction (Ω_1), which is defined as 1:C:H1; many-to-many translation pair extraction from connected existing edges (Ω_2), which is defined as 1:M:H1; and many-to-many translation pair extraction from connected existing and new edges (Ω_3), which is defined as 2:M:H1. We also use the inverse consultation method (IC) and translation pairs generated from Cartesian product of input dictionaries (CP) as baselines.

6.1 Experimental Settings

We have four case studies; one of the closely related low-resource languages of the Austronesian language family and three of the high-resource Indo-European languages. The language similarities shown in Table 4 were computed using ASJP. We generate translation pairs from the Cartesian Product within and across transgraph to be used in the evaluation as shown in Figure 12.

We selected the Indonesian ethnic languages Minangkabau (min) and Riau Mainland Malay (zlm) with the Indonesian language (ind) as the pivot for our first case study (min-ind-zlm). Even though Malaysian Malay (zlm) is not part of Indonesian ethnic languages, it is very similar to Riau Mainland Malay. In fact, Riau Mainland Malay is one of the Malaysian Malay dialects [27]. Since there is no available machine-readable dictionary of Indonesian to Riau Mainland Malay, we used the available machine-readable dictionary of Indonesian to Malaysian Malay (zlm) for the case study min-ind-zlm. A trilingual Indonesian, Malaysian Malay, and Riau Mainland Malay

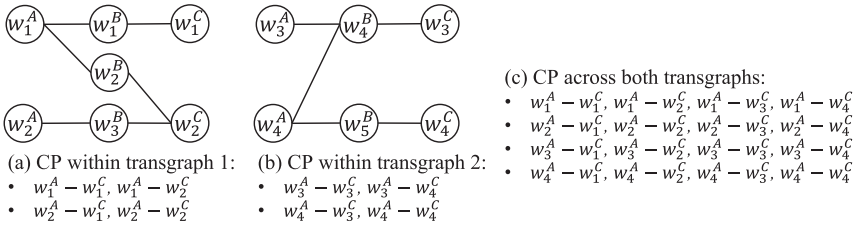


Fig. 12. Example of extracting translation pair candidates from Cartesian Product (CP).

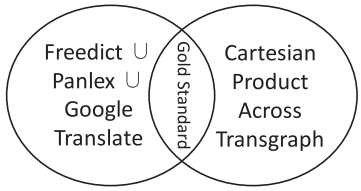


Fig. 13. Creating the gold standard for the high-resource case studies.

Table 5. Dictionaries for Evaluation

Source	Number of Translation		
Freedict	deu-nld ∪ nld-deu = 35,962	spa-por = 333	deu-ita ∪ ita-deu = 6,152
Panlex	deu-nld = 405,076	spa-por = 343,665	deu-ita = 475,461
Google Translate*	deu-nld ∪ nld-deu = 1,924	spa-por ∪ por-spa = 1,338	deu-ita ∪ ita-deu = 1,790
TOTAL	deu-nld = 406,370	spa-por = 344,126	deu-ita = 476,172

**Translating all headwords from CP within the transgraphs.*

speaker thoroughly cleansed the dictionary by deleting or editing Malaysian Malay words that are not present in the Riau Mainland Malay language. We generate full-matching translation pairs (Cartesian product within transgraph from input dictionaries), verified by the Minangkabau-Malay bilingual speaker via crowdsourcing, and took them as the gold standard for calculating precision and recall.

The Proto-Indo-European language is the common ancestor of the Indo-European language family from which the rest of our case-study languages originate. The second case study (deu-eng-nld) targets high-resource languages of German (deu) and Dutch (nld) with English (eng) as the pivot. The third case study (spa-eng-por) uses the Spanish (spa) and Portugese (por) languages with English (eng) as the pivot. The fourth case study (deu-eng-ita) uses the German (deu) and Italian (ita) languages with English (eng) as the pivot. We utilize Freedict, an open source online bilingual dictionary database,[1] as input dictionaries and combination of Freedict, Panlex, another bilingual dictionary databases,[2] and Google Translate[3] as shown in Table 5 as dictionaries for evaluation to create a gold standard. We use Google Translate to translate all headwords from Cartesian Product (CP) within the transgraphs. The gold standard is obtained by intersecting the combination of dictionaries for evaluation with CP across transgraph as shown in Figure 13. The structure of the input dictionaries and the gold standard for every case studies can be found

[1]http://freedict.org.
[2]http://panlex.org.
[3]http://translate.google.com.

Table 6. Structure of Input Dictionaries and Gold Standard

Case Study	min-ind-zlm			deu-eng-nld			spa-eng-por			deu-eng-ita		
Language	min	ind	zlm	deu	eng	nld	spa	eng	por	deu	eng	ita
Headword	520	625	681	968	673	1,183	600	849	986	1,157	1,340	842
CP within transgraph			1,757			5,790			2,526			2,959
CP across transgraph			354,120			1,145,144			591,600			974,194
Gold Standard			1,246			1,438			1,069			1,503

Table 7. Translation Relationship of Input Dictionaries

Case Study	Bilingual Dictionary	Translation Relationship							
		1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8
min-ind-zlm	min-ind	267	210	36	5	1	1	0	0
	zlm-ind	563	115	3	0	0	0	0	0
deu-eng-nld	deu-eng	785	165	16	2	0	0	0	0
	nld-eng	705	410	49	14	3	1	1	0
spa-eng-por	spa-eng	204	289	86	16	2	2	1	0
	por-eng	458	370	116	33	7	2	0	0
deu-eng-ita	deu-eng	971	154	30	2	0	0	0	0
	ita-eng	256	421	129	25	7	2	1	1

Table 8. Size of the Biggest Transgraph

Case Study ($L_1 - P - L_2$)	L_1 Words	P Words	L_2 Words	Edges
min-ind-zlm	8	14	18	39
deu-eng-nld	4,669	2,486	6,864	18,548
spa-eng-por	2,347	2,465	4,460	15,043
deu-eng-ita	650	822	597	2,242

in Table 6. The translation relationship of the input dictionaries varies from one-to-one until one-to-eight as shown in Table 7. For the low-resource case study, that is, min-ind-zlm, the input dictionaries only have few one-to-many translation relations compared to the high-resource case studies. This shows that there are many potential missing senses in the input dictionaries. Consequently, sometimes we can miss some translation pair candidates across the transgraphs. Therefore, in this article, we limit our scope to extracting translation pairs within the transgraphs.

We do not discriminate both single-word and multi-words expressions in the input dictionaries. After constructing the transgraphs from the input dictionaries, we find one big transgraph for each high-resource language case study as shown in Table 8. Sometimes, for high-resource languages where the input dictionaries have many shared meanings via the pivot words, a big transgraph can be generated that potentially leads to a computational complexity when we formalize and solve it. Nevertheless, for a low-resource language where we can expect that the input dictionaries only have a few shared meanings via the pivot words, the size of the transgraph is feasible to be formalized and solved. Therefore, for the sake of simplicity, we ignore any big transgraphs in these experiments.

Different users are likely to have different motivations, priorities, and preferences when creating a bilingual dictionary. For high-resource languages, some users tend to prioritize precision over recall while for low-resource languages, most users tend to prioritize recall to enrich the language resource. In this article, we optimize the hyperparameters (cognate threshold and cognate

synonym threshold) with a grid search by incrementing the cognate threshold from 0 to the highest cost of violating the constraints with 0.01 intervals and incrementing the cognate synonym threshold from 0 to 1 with 0.01 intervals to find the highest F-score.

6.2 Experiment Result

In all experiments and all case studies, all transgraphs are fully symmetrically connected on the third cycle, and thus all possible translation pair candidates are reached. To extract many-to-many translation pairs, in the first step, that is, cognate recognition and the second step, that is, cognate synonym recognition, the soft-constraint violation threshold is set to reject all translation pairs returned by SATSolver that incurred a higher cost than the cognate threshold and cognate synonym threshold as shown in Algorithm 2 line number 4 and 10, respectively. Even though using the threshold to prioritize precision could yield the highest precision, the recall can be very low. Similarly, even though using the threshold to prioritize recall could yield the highest recall, the precision can also be harmed. Blindly prioritizing the precision over the recall or recall over the precision might not be a good strategy when implementing the framework.

6.2.1 Threshold Yielding the Highest F-score. To obtain a good strategy when we want to implement the framework, a balance between precision and recall is crucial. We calculate a harmonic mean of precision and recall using the traditional F1-measure or balanced F1-score by weighting the precision and recall equally. Based on user preference and priority, F0.5-score can be used when precision is considered more important, and F2-score can be used when recall is preferred. The results of all four case studies that targeted the threshold yielding the highest F-score are shown in Table 9. For the case study min-ind-zlm, our best yielding M-M result method (2:S:H14) yields 0.4% higher F-score than our previous best yielding M-M result method (2:M:H1), 3.4% higher F-score than CP, and 12.9 times higher F-score than IC, while our best yielding 1-1 result method (3:C:H14) yields 1.3% higher precision than our previous method (1:C:H1). The high F-score of the CP in the case study min-ind-zlm indicates how very closely related the input languages are. For the case study deu-eng-nld, our best yielding M-M result method (1:S:H124) yields 0.2% higher F-score than our previous best yielding M-M result method (1:M:H1), 46% higher F-score than CP, and 2.9 times higher F-score than IC, while our best yielding 1-1 result method (3:C:H14) yields 5.5% higher precision than our previous method (1:C:H1). For the case study spa-eng-por, our best yielding M-M result method (1:S:H14) yields 0.6% higher F-score than our previous best yielding M-M result method (1:M:H1), 26.3% higher F-score than CP, and 27.3% higher F-score than IC, while our best yielding 1-1 result method (3:C:H34) yields 3.6% higher precision than our previous method (1:C:H1). For the case study deu-eng-ita, our best yielding M-M result method (1:S:H14) yields 0.2% higher F-score than our previous best yielding M-M result method (1:M:H1), 30.7% higher F-score than CP, and 3.2 times higher F-score than IC, while our best yielding 1-1 result method (3:C:H134) yields 3.6% higher precision than our previous method (1:C:H1).

To enrich the bilingual dictionary result for low-resource languages, cognates and cognate synonyms recognition with higher cycles is the best approach. The exact number of cycles can be customized based on the priority and preference as regards the precision-recall tradeoff. The cognates and cognate synonyms recognition with one-cycle is recommended for attaining the highest F-score result, since for almost all case studies in our experiments except min-ind-zlm, it always realized the highest F-score.

For the case study deu-eng-nld, the best one-to-one cognate (3:C:H14) method precision is unexpectedly low, 0.474, while the lower language similarity case studies (spa-eng-por and deu-eng-ita) with the same cycle have higher precision (0.716 and 0.621, respectively). The case study deu-eng-nld always yielded lower F-scores than case studies deu-eng-ita and spa-eng-por when the methods

Table 9. Threshold Yielding the Highest F-score

Case Study	Method	Cognate Threshold	Cognate Synonym Threshold	Precision	Recall	F-score
	3:S:H14 (M-M)	4.79	1	0.656	0.998	0.792
	2:S:H14 (M-M)	4.79	0.74	0.735	0.923	0.818
	1:S:H14 (M-M)	4.17	1	0.836	0.713	0.770
	3:C:H14 (1-1)	4.79		0.884	0.331	0.481
	2:C:H14 (1-1)	4.79		0.884	0.331	0.481
min-ind-zlm	1:C:H14 (1-1)	4.17		0.878	0.328	0.478
	Baseline: 2:M:H1 (M-M)			0.713	0.953	0.815
	Baseline: 1:M:H1 (M-M)			0.836	0.713	0.770
	Baseline: 1:C:H1 (1-1)			0.873	0.327	0.475
	Baseline: CP (M-M)			0.654	0.998	0.791
	Baseline: IC (M-M)			0.950	0.031	0.059
	3:S:H14 (M-M)	1.97	1	0.230	0.926	0.368
	2:S:H14 (M-M)	1.97	0.49	0.323	0.707	0.443
	1:S:H124 (M-M)	4.1	0.99	0.400	0.820	0.537
	3:C:H14 (1-1)	1.97		0.474	0.250	0.328
	2:C:H14 (1-1)	1.97		0.474	0.250	0.328
deu-eng-nld	1:C:H124 (1-1)	4.1		0.472	0.249	0.327
	Baseline: 2:M:H1 (M-M)			0.257	0.919	0.402
	Baseline: 1:M:H1 (M-M)			0.397	0.821	0.536
	Baseline: 1:C:H1 (1-1)			0.447	0.238	0.311
	Baseline: CP (M-M)			0.230	0.926	0.368
	Baseline: IC (M-M)			0.612	0.078	0.138
	3:S:H34 (M-M)	3.01	1	0.368	0.870	0.517
	2:S:H34 (M-M)	3.01	0.49	0.467	0.751	0.576
	1:S:H14 (M-M)	3.21	0.66	0.569	0.765	0.653
	3:C:H34 (1-1)	3.01		0.716	0.367	0.486
	2:C:H34 (1-1)	3.01		0.716	0.367	0.486
spa-eng-por	1:C:H14 (1-1)	3.21		0.717	0.367	0.486
	Baseline: 2:M:H1 (M-M)			0.389	0.870	0.537
	Baseline: 1:M:H1 (M-M)			0.538	0.818	0.649
	Baseline: 1:C:H1 (1-1)			0.695	0.356	0.471
	Baseline: CP (M-M)			0.368	0.870	0.517
	Baseline: IC (M-M)			0.708	0.402	0.513
	3:S:H134 (M-M)	6.14	1	0.320	0.630	0.425
	2:S:H134 (M-M)	6.14	0.85	0.477	0.534	0.504
	1:S:H14 (M-M)	6.14	0.85	0.544	0.564	0.554
	3:C:H134 (1-1)	6.14		0.621	0.310	0.413
	2:C:H134 (1-1)	6.14		0.621	0.310	0.413
deu-eng-ita	1:C:H14 (1-1)	6.14		0.626	0.310	0.415
	Baseline: 2:M:H1 (M-M)			0.377	0.627	0.471
	Baseline: 1:M:H1 (M-M)			0.542	0.565	0.553
	Baseline: 1:C:H1 (1-1)			0.600	0.298	0.398
	Baseline: CP (M-M)			0.320	0.630	0.424
	Baseline: IC (M-M)			0.930	0.071	0.131

⟨*situatedMethod*⟩ ::= ⟨*cycle*⟩" : "⟨*method*⟩" : "⟨*heuristic*⟩ where *cycle*: symmetry assumption cycle where cycle ≥ 1, *method*: *C* as a cognate recognition ($CNF_{cognate}$) or *S* as a cognate and cognate synonym recognition ($CNF_{cognate}$ + $CNF_{cognateSynonym}$) or *M* as a many-to-many approach (Ω_2 and Ω_3) in our previous work [20], *heuristic*: an individual or combined heuristics where H1234 means a combination of heuristic 1 (cognate pair coexistence probability), heuristic 2 (missing contribution rate toward cognate pair coexistence), heuristic 3 (polysemy pivot ambiguity rate), and heuristic 4 (cognate form similarity). CP: Cartesian Product; IC: Inverse Consultation [31]; 1-1: one-to-one translation pair results; M-M: many-to-many translation pair results;

that generate many-to-many results were applied. We believe that inadequacy of the gold standard was the cause of this counterintuitive result. For the case study deu-eng-nld, if we look at the ratio of the size of the Cartesian product across transgraph in Table 6 and the size of the combined dictionaries for evaluation in Table 5, relative to the ratio of the gold standard and the Cartesian product within the transgraph, it is obvious that the ratio is inadequate compared to the other case study languages.

Table 10. Comparison of the Proposed Methods and the Previous Method: Case study min-ind-zlm

		Previous Method			Proposed Method					
Comparison	Transgraph	Precision	Recall	F-score	Precision	Diff.	Recall	Diff.	F-score	Diff.
	0-24	0.92	0.548	0.687	0.92	0	0.548	0	0.687	0
	25-40	0.813	0.542	0.65	0.813	0	0.542	0	0.65	0
	41-56	0.813	0.52	0.634	0.875	+0.063	0.56	+0.04	0.683	+0.049
	57-72	1	0.516	0.681	1	0	0.516	0	0.681	0
	73-88	0.9	0.621	0.735	0.9	0	0.621	0	0.735	0
	89-104	0.889	0.471	0.615	0.889	0	0.471	0	0.615	0
	105-120	0.63	0.447	0.523	0.667	+0.037	0.474	+0.026	0.554	+0.031
1-1*	121-136	0.552	0.533	0.542	0.552	0	0.533	0	0.542	0
	137-152	0.828	0.5	0.623	0.862	+0.034	0.521	+0.021	0.649	+0.026
	153-168	0.966	0.346	0.509	1	+0.034	0.358	+0.012	0.527	+0.018
	169-184	1	0.352	0.52	1	0	0.352	0	0.52	0
	185-200	1	0.34	0.508	1	0	0.34	0	0.508	0
	201-216	0.975	0.312	0.473	0.975	0	0.312	0	0.473	0
	217-232	0.889	0.294	0.442	0.889	0	0.294	0	0.442	0
	233-248	0.866	0.179	0.296	0.878	+0.012	0.181	+0.003	0.301	+0.004
	0-24	0.913	1	0.955	0.913	0	1	0	0.955	0
	25-40	0.75	1	0.857	0.75	0	1	0	0.857	0
	41-56	0.781	1	0.877	0.781	0	1	0	0.877	0
	57-72	0.969	1	0.984	0.969	0	1	0	0.984	0
	73-88	0.725	1	0.841	0.725	0	1	0	0.841	0
	89-104	0.85	1	0.919	0.864	+0.014	1	0	0.927	+0.008
	105-120	0.644	1	0.784	0.644	0	1	0	0.784	0
M-M**	121-136	0.492	1	0.659	0.492	0	1	0	0.659	0
	137-152	0.774	1	0.873	0.774	0	1	0	0.873	0
	153-168	0.92	1	0.959	0.92	0	1	0	0.959	0
	169-184	0.938	1	0.968	0.938	0	1	0	0.968	0
	185-200	0.906	0.99	0.946	0.906	0	0.99	0	0.946	0
	201-216	0.886	0.992	0.936	0.886	0	0.992	0	0.936	0
	217-232	0.744	0.985	0.848	0.772	+0.028	0.949	−0.037	0.851	+0.003
	233-248	0.544	0.864	0.667	0.544	0	0.864	0	0.667	0

*Comparison between the previous method (1:C:H1 / Ω_1) [20] and the proposed method (2:C:H14) that yields one-to-one translation pair results.

**Comparison between the previous method (2:M:H1 / Ω_3) [20] and the proposed method (2:S:H14) that yields many-to-many translation pair results.

6.2.2 Statistical Significant Test. To show that our proposed methods are statistically significant compared to our previous methods [20], as listed in Tableless 10–13, for each case study, first, we split the dataset into several datapoints (transgraphs), and then we compare the potentially best methods yielding the most many-to-many translation pairs (M-M), that is, the 2:S:H14 to our previous method that potentially yielding the most many-to-many translation pairs (M-M), that is, 2:M:H1. We also compare the potentially best methods yielding the most one-to-one translation pairs (1-1), that is, the 2:C:H14 to our previous method that yielding one-to-one translation pairs (1-1), that is, 1:C:H1. Student's paired t-test is a good statistical procedure used in Information Retrieval research to determine whether the mean difference between two sets of observations is

Table 11. Comparison of the Proposed Methods and the Previous Method: Case study deu-eng-nld

		Previous Method			Proposed Method					
Comparison	Transgraph	Precision	Recall	F-score	Precision	Diff.	Recall	Diff.	F-score	Diff.
	0-16	0.529	0.9	0.667	0.588	+0.059	1	+0.1	0.741	+0.074
	17-34	0.478	0.478	0.478	0.522	+0.043	0.522	+0.043	0.522	+0.043
	35-52	0.594	0.463	0.521	0.594	0	0.463	0	0.521	0
	53-70	0.286	0.25	0.267	0.286	0	0.25	0	0.267	0
	71-88	0.406	0.271	0.325	0.5	+0.094	0.333	+0.063	0.4	+0.075
	89-106	0.447	0.333	0.382	0.447	0	0.333	0	0.382	0
	107-124	0.641	0.321	0.427	0.667	+0.026	0.333	+0.013	0.444	+0.017
1-1*	125-142	0.455	0.235	0.31	0.455	0	0.235	0	0.31	0
	143-160	0.439	0.22	0.293	0.512	+0.073	0.256	+0.037	0.341	+0.049
	161-178	0.333	0.237	0.277	0.426	+0.093	0.303	+0.066	0.354	+0.077
	179-196	0.526	0.265	0.353	0.517	−0.009	0.265	0	0.351	−0.002
	197-214	0.435	0.195	0.269	0.435	0	0.195	0	0.269	0
	215-232	0.408	0.228	0.293	0.38	−0.028	0.213	−0.016	0.273	−0.02
	233-250	0.41	0.211	0.279	0.457	+0.047	0.23	+0.019	0.306	+0.027
	251-268	0.446	0.135	0.208	0.485	+0.039	0.14	+0.004	0.217	+0.009
	0-16	0.417	1	0.588	0.417	0	1	0	0.588	0
	17-34	0.435	0.87	0.58	0.455	+0.02	0.87	0	0.597	+0.017
	35-52	0.559	0.927	0.697	0.587	+0.028	0.902	−0.024	0.712	+0.014
	53-70	0.329	0.844	0.474	0.329	0	0.844	0	0.474	0
	71-88	0.392	0.833	0.533	0.392	0	0.833	0	0.533	0
	89-106	0.349	0.882	0.5	0.366	+0.017	0.804	−0.078	0.503	+0.003
	107-124	0.531	0.987	0.691	0.531	0	0.987	0	0.691	0
M-M**	125-142	0.363	0.729	0.484	0.389	+0.026	0.659	−0.071	0.489	+0.005
	143-160	0.371	0.915	0.528	0.371	0	0.915	0	0.528	0
	161-178	0.274	0.961	0.427	0.304	+0.029	0.763	−0.197	0.434	+0.008
	179-196	0.33	0.947	0.49	0.33	0	0.947	0	0.49	0
	197-214	0.254	0.675	0.369	0.287	+0.033	0.643	−0.032	0.397	+0.027
	215-232	0.224	0.898	0.358	0.271	+0.047	0.646	−0.252	0.381	+0.023
	233-250	0.197	0.87	0.322	0.254	+0.056	0.671	−0.199	0.368	+0.046
	251-268	0.199	0.849	0.323	0.301	+0.102	0.561	−0.288	0.392	+0.069

*Comparison between the previous method (1:C:H1 / Ω_1) [20] and the proposed method (2:C:H14) that yields one-to-one translation pair results.

**Comparison between the previous method (2:M:H1 / Ω_3) [20] and the proposed method (2:S:H14) that yields many-to-many translation pair results.

zero [28]. It is very useful to show that our proposed methods are truly better than our previous methods rather than performed better by chance. In a student's paired t-test, each subject or entity is measured twice, resulting in pairs of observations. In this article, we use the same set of datapoints and conduct the student's paired t-test with precision and F-score as measures. Since we expect that our proposed methods have improvement compared to our previous methods, we choose a one-tailed t-test. There are two sets of null hypotheses (precision null hypotheses and F-score null hypotheses), which are that the true precision or F-score means difference between the proposed methods and our previous methods are equal to zero. We decide 0.05 cutoff value for determining statistical significance that corresponds to a 5% (or less) chance of obtaining a result like

Table 12. Comparison of the Proposed Methods and the Previous Method: Case Study spa-eng-por

		Previous Method			Proposed Method					
Comparison	Transgraph	Precision	Recall	F-score	Precision	Diff.	Recall	Diff.	F-score	Diff.
	0-24	1	0.833	0.909	1	0	0.833	0	0.909	0
	25-45	0.714	0.75	0.732	0.714	0	0.75	0	0.732	0
	46-66	0.81	0.68	0.739	0.81	0	0.68	0	0.739	0
	67-87	0.762	0.421	0.542	0.762	0	0.421	0	0.542	0
	88-108	0.667	0.467	0.549	0.714	+0.048	0.5	+0.033	0.588	+0.039
	109-129	0.762	0.471	0.582	0.81	+0.048	0.5	+0.029	0.618	+0.036
	130-150	0.64	0.364	0.464	0.68	+0.04	0.386	+0.023	0.493	+0.029
1-1*	151-171	0.724	0.382	0.5	0.69	−0.034	0.364	−0.018	0.476	−0.024
	172-192	0.9	0.351	0.505	0.9	0	0.351	0	0.505	0
	193-213	0.6	0.296	0.397	0.6	0	0.296	0	0.397	0
	214-234	0.61	0.212	0.314	0.585	−0.024	0.203	−0.008	0.302	−0.013
	235-255	0.587	0.287	0.386	0.63	+0.043	0.309	+0.021	0.414	+0.029
	256-276	0.577	0.288	0.385	0.615	+0.038	0.308	+0.019	0.41	+0.026
	277-297	0.678	0.276	0.392	0.712	+0.034	0.29	+0.014	0.412	+0.02
	298-318	0.708	0.221	0.337	0.74	+0.031	0.231	+0.01	0.352	+0.015
	0-24	1	0.833	0.909	1	0	0.833	0	0.909	0
	25-45	0.714	0.75	0.732	0.714	0	0.75	0	0.732	0
	46-66	0.75	0.84	0.792	0.75	0	0.84	0	0.792	0
	67-87	0.667	0.737	0.7	0.667	0	0.737	0	0.7	0
	88-108	0.585	0.8	0.676	0.585	0	0.8	0	0.676	0
	109-129	0.596	0.824	0.691	0.596	0	0.824	0	0.691	0
	130-150	0.667	0.909	0.769	0.678	+0.011	0.909	0	0.777	+0.007
M-M**	151-171	0.632	0.782	0.699	0.646	+0.014	0.764	−0.018	0.7	+0.001
	172-192	0.663	0.766	0.711	0.663	0	0.766	0	0.711	0
	193-213	0.46	0.704	0.556	0.46	0	0.704	0	0.556	0
	214-234	0.438	0.534	0.481	0.458	+0.021	0.508	−0.025	0.482	+0.001
	235-255	0.433	0.83	0.569	0.433	0	0.83	0	0.569	0
	256-276	0.359	0.817	0.499	0.359	0	0.817	0	0.499	0
	277-297	0.36	0.862	0.508	0.433	+0.073	0.697	−0.166	0.534	+0.026
	298-318	0.255	0.779	0.384	0.359	+0.104	0.59	−0.189	0.446	+0.062

*Comparison between the previous method (1:C:H1 / Ω_1) [20] and the proposed method (2:C:H14) that yields one-to-one translation pair results.

**Comparison between the previous method (2:M:H1 / Ω_3) [20] and the proposed method (2:S:H14) that yields many-to-many translation pair results.

the one that was observed if the null hypotheses were true. For all case studies min-ind-zlm, deu-eng-nld, spa-eng-por, and deu-eng-ita, we reject the precision null hypotheses, since the p-value of the tests are 0.00732, 0.00007, 0.00398, 0.00464, respectively, which are all smaller than 0.05. For all case studies min-ind-zlm, deu-eng-nld, spa-eng-por, and deu-eng-ita, we also reject the F-score null hypotheses, since the p-value of the tests are 0.01673, 0.00034, 0.00652, and 0.00783, respectively, which are all smaller than 0.05. Thus, our proposed methods have statistically significant improvement of precision and F-score compared to our previous methods.

6.2.3 Hyperparameter Optimization. We have shown that our methods outperformed the baselines in the previous sections. Nevertheless, before implementing our model in a big scale, we

Table 13. Comparison of the Proposed Methods and the Previous Method: Case study deu-eng-ita

		Previous Method			Proposed Method					
Comparison	Transgraph	Precision	Recall	F-score	Precision	Diff.	Recall	Diff.	F-score	Diff.
1-1*	0-34	0.943	0.367	0.528	0.943	0	0.367	0	0.528	0
	35-64	0.633	0.235	0.342	0.7	+0.067	0.259	+0.025	0.378	+0.036
	65-94	0.7	0.3	0.42	0.667	−0.033	0.286	−0.014	0.4	−0.02
	95-124	0.533	0.246	0.337	0.667	+0.133	0.308	+0.062	0.421	+0.084
	125-154	0.667	0.274	0.388	0.667	0	0.274	0	0.388	0
	155-184	0.5	0.167	0.25	0.533	+0.033	0.178	+0.011	0.267	+0.017
	185-214	0.567	0.23	0.327	0.633	+0.067	0.257	+0.027	0.365	+0.038
	215-244	0.646	0.316	0.425	0.667	+0.021	0.327	+0.01	0.438	+0.014
	245-274	0.694	0.256	0.374	0.673	−0.02	0.248	−0.008	0.363	−0.011
	275-304	0.689	0.341	0.456	0.711	+0.022	0.352	+0.011	0.471	+0.015
	305-334	0.556	0.197	0.291	0.587	+0.031	0.213	+0.016	0.312	+0.021
	335-364	0.561	0.182	0.275	0.542	−0.019	0.182	0	0.272	−0.002
	365-394	0.54	0.177	0.267	0.556	+0.016	0.182	+0.005	0.275	+0.008
	395-424	0.519	0.169	0.256	0.532	+0.013	0.174	+0.004	0.262	+0.006
	425-454	0.544	0.184	0.275	0.562	+0.017	0.189	+0.005	0.283	+0.007
M-M**	0-34	0.946	0.389	0.551	0.946	0	0.389	0	0.551	0
	35-64	0.672	0.481	0.561	0.672	0	0.481	0	0.561	0
	65-94	0.627	0.529	0.574	0.627	0	0.529	0	0.574	0
	95-124	0.593	0.538	0.565	0.593	0	0.538	0	0.565	0
	125-154	0.61	0.493	0.545	0.61	0	0.493	0	0.545	0
	155-184	0.583	0.389	0.467	0.583	0	0.389	0	0.467	0
	185-214	0.633	0.514	0.567	0.633	0	0.514	0	0.567	0
	215-244	0.515	0.52	0.518	0.515	0	0.52	0	0.518	0
	245-274	0.535	0.406	0.462	0.535	0	0.406	0	0.462	0
	275-304	0.455	0.549	0.498	0.455	0	0.549	0	0.498	0
	305-334	0.444	0.441	0.443	0.444	0	0.441	0	0.443	0
	335-364	0.407	0.409	0.408	0.419	+0.012	0.398	−0.011	0.408	0
	365-394	0.367	0.438	0.399	0.401	+0.034	0.422	−0.016	0.411	+0.012
	395-424	0.331	0.462	0.386	0.379	+0.048	0.419	−0.042	0.398	+0.013
	425-454	0.226	0.488	0.309	0.339	+0.112	0.336	−0.152	0.338	+0.028

*Comparison between the previous method (1:C:H1 / Ω_1) [20] and the proposed method (2:C:H14) that yields one-to-one translation pair results.

**Comparison between the previous method (2:M:H1 / Ω_3) [20] and the proposed method (2:S:H14) that yields many-to-many translation pair results.

need to validate how good our model perform in practice with unknown data. Since there are not enough data available to partition it into separate training and test sets without losing significant modelling or testing capability, a good way to properly estimate model prediction performance is to use cross-validation as a powerful general technique. Due to the computational complexity of our model, we conduct threefold cross-validation to predict the optimal hyperparameters (cognate threshold and cognate synonym threshold) to gain the highest F-score as shown in Table 14. We optimize the hyperparameters with a grid search by incrementing the cognate threshold from 0 to the highest cost of violating the constraints with 0.01 intervals and incrementing the cognate synonym threshold from 0 to 1 with 0.01 intervals to find the highest F-score. We choose the same

Table 14. Cognate Threshold and Cognate Synonym Threshold Optimization

			Optimal Threshold		Testing on Unknown Data				
Case Study	Method	Validation Set	Cognate	Cognate Synonym	Test Set	Precision	Recall	F-score	Mean F-score
min-ind-zlm	2CH14	0-82, 83-165	1.35	-	166-248	0.933	0.257	0.403	
		0-82, 166-248	4.79	-	83-165	0.786	0.471	0.589	0.559
		83-165, 166-248	4.79	-	0-82	0.916	0.547	0.685	
	2SH14	0-82, 83-165	1.99	1	166-248	0.688	0.933	0.792	
		0-82, 166-248	4.79	0.26	83-165	0.729	1	0.843	0.853
		83-165, 166-248	4.79	0.26	0-82	0.858	1	0.924	
deu-eng-nld	2CH14	0-90, 91-179	1.85	-	180-268	0.467	0.185	0.265	
		0-90, 180-268	1.97	-	91-179	0.493	0.285	0.361	0.359
		91-179, 180-268	1.97	-	0-90	0.485	0.423	0.452	
	2SH14	0-90, 91-179	1.85	1	180-268	0.219	0.86	0.35	
		0-90, 180-268	1.97	0.51	91-179	0.361	0.893	0.514	0.474
		91-179, 180-268	1.97	0.51	0-90	0.41	0.878	0.559	
spa-eng-por	2CH14	0-106, 107-212	2.96	-	213-318	0.676	0.268	0.384	
		0-106, 213-318	3.21	-	107-212	0.724	0.366	0.486	0.525
		107-212, 213-318	3.21	-	0-106	0.804	0.628	0.705	
	2SH14	0-106, 107-212	2.96	0.51	213-318	0.394	0.636	0.487	
		0-106, 213-318	3.21	0.51	107-212	0.603	0.756	0.671	0.639
		107-212, 213-318	3.21	0.51	0-106	0.719	0.803	0.759	
deu-eng-ita	2CH14	0-150, 151-302	1.5	-	303-454	0.557	0.202	0.297	
		0-150, 303-454	6.14	-	151-302	0.652	0.279	0.391	0.371
		151-302, 303-454	6.14	-	0-150	0.735	0.3	0.426	
	2SH14	0-150, 151-302	1.5	0.01	303-454	0.341	0.414	0.374	
		0-150, 303-454	6.14	0.56	151-302	0.531	0.481	0.505	0.479
		151-302, 303-454	6.14	0.56	0-150	0.67	0.478	0.558	

methods as in Tables 10–13, the potentially best methods yielding the most one-to-one translation pairs (1-1), that is, the 2:C:H14 and the potentially best methods yielding the most many-to-many translation pairs (M-M), that is, the 2:S:H14. For all case studies, the mean F-score approaches the mean F-score of the overfitting model in Tables 10–13.

7 CONCLUSION

Our strategy to create high-quality many-to-many translation pairs between closely related languages consists of two steps. We first recognize cognates from direct and indirect connectivity via pivot word(s) by iterating multiple symmetry assumption cycles to reach more cognates in the transgraph. Once we obtain a list of cognates, the next step identifies synonyms of those cognates.

The result of case studies showed that our method offers good performance on weakly related high-resource languages. Thus, our method has the potential to complement other bilingual dictionary creation methods like word alignment models using parallel corpora. Our method shows particularly high performance on the closely related low-resource language case study. Our proposed methods have statistically significant improvement of precision and F-score compared to our previous methods in spite of sacrificing the recall a little bit.

Our key research contribution is a generalized constraint-based bilingual lexicon induction framework for closely related low-resource languages. This generalization makes our method applicable for a wider range of language groups than the one-to-one approach. Our customizable approach allows the user to conduct cross validation to predict the optimal hyperparameters (cognate threshold and cognate synonym threshold) with various combination of heuristics and number of symmetry assumption cycles to gain the highest F-score. To the best of our knowledge, our study is the first attempt to recognize both cognates and cognate synonyms in bilingual lexicon induction.

REFERENCES

[1] Carlos Ansótegui, María Luisa Bonet, and Jordi Levy. 2009. Solving (weighted) partial MaxSAT through satisfiability testing. In *Theory and Applications of Satisfiability Testing-SAT 2009*. Springer, 427–440.

[2] Armin Biere, Marijn Heule, and Hans van Maaren. 2009. *Handbook of Satisfiability*. Vol. 185. IOS Press.

[3] Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Comput. Ling.* 16, 2 (1990), 79–85.

[4] Lyle Campbell. 2013. *Historical Linguistics*. Edinburgh University Press.

[5] Lyle Campbell and William J. Poser. 2008. Language classification. *History and Method*. Cambridge University Press, Cambridge (2008).

[6] Hervé Déjean, Éric Gaussier, and Fatia Sadat. 2002. An approach based on multilingual thesauri and model combination for bilingual lexicon extraction. In *Proceedings of the 19th International Conference on Computational Linguistics, Vol. 1 (COLING'02)*. Association for Computational Linguistics, Stroudsburg, PA, 1–7. DOI: http://dx.doi.org/10.3115/1072228.1072394

[7] Pascale Fung. 1995. Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus. In *Proceedings of the 3rd Workshop on Very Large Corpora*. 173–183.

[8] Pascale Fung. 1998. A statistical view on bilingual lexicon extraction: From parallel corpora to non-parallel corpora. In *Machine Translation and the Information Soup*. Springer, 1–17.

[9] Charlotte Gooskens. 2006. Linguistic and extra-linguistic predictors of inter-scandinavian intelligibility. *Ling. Netherlands* 23, 1 (2006), 101–113.

[10] Ahlem Ben Hassine, Shigeo Matsubara, and Toru Ishida. 2006. A constraint-based approach to horizontal web service composition. In *Proceedings of the International Semantic Web Conference*. Springer, 130–143.

[11] Eric W. Holman, Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, and others. 2011. Automated dating of the world's language families based on lexical similarity. *Curr. Anthropol.* 52, 6 (2011), 841–875.

[12] Toru Ishida. 2011. *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*. Springer.

[13] Winfred P. Lehmann. 2013. *Historical Linguistics: An Introduction*. Routledge.

[14] M. Paul Lewis, Gary F. Simons, and Charles D. Fennig (Eds.). 2015. *Ethnologue: Languages of the World (18th ed.)*. SIL International, Dallas, TX.

[15] Gideon S. Mann and David Yarowsky. 2001. Multipath translation lexicon induction via bridge languages. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*. Association for Computational Linguistics, 1–8.

[16] Jun Matsuno and Toru Ishida. 2011. Constraint optimization approach to context based word selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'11)*, Vol. 22.

[17] I. Dan Melamed. 1995. Automatic evaluation and uniform filter cascades for inducing N-best translation lexicons. *CoRR* cmp-lg/9505044 (1995). Retrieved from http://arxiv.org/abs/cmp-lg/9505044.

[18] Preslav Nakov and Hwee Tou Ng. 2012. Improving statistical machine translation for a resource-poor language using related resource-rich languages. *J. Artif. Intell. Res.* 44 (2012), 179–222.

[19] Preslav Nakov and Jörg Tiedemann. 2012. Combining word-level and character-level models for machine translation between closely related languages. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Volume 2*. Association for Computational Linguistics, 301–305.

[20] Arbi Haza Nasution, Yohei Murakami, and Toru Ishida. 2016. Constraint-based bilingual lexicon induction for closely related languages. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*. Paris, France, 3291–3298.

[21] Reinhard Rapp. 1995. Identifying word translations in non-parallel texts. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 320–322.

[22] John Richardson, Toshiaki Nakazawa, and Sadao Kurohashi. 2015. Pivot-based topic models for low-resource lexicon extraction. In *Proceedings of the Pacific Asia Conference on Language, Information and Computation (PACLIC'15)*.

[23] C. J. Van Rijsbergen. 1979. *Information Retrieval* (2nd ed.). Butterworth-Heinemann, Newton, MA.

[24] Xabier Saralegi, Iker Manterola, and Inaki San Vicente. 2011. Analyzing methods for improving precision of pivot based bilingual dictionaries. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 846–856.

[25] Kevin P. Scannell. 2006. Machine translation for closely related language pairs. In *Proceedings of the Workshop Strategies for Developing Machine Translation for Minority Languages*. Citeseer, 103–109.

[26] Lloyd S. Shapley. 1953. A value for n-person games. *Contrib. Theor. Games* 2, 28 (1953), 307–317.

[27] Gary F. Simons and Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the World*, 20th ed. (2017).

[28] Mark D. Smucker, James Allan, and Ben Carterette. 2007. A comparison of statistical significance tests for information retrieval evaluation. In *Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management (CIKM'07)*. ACM, New York, NY, 623–632. DOI: http://dx.doi.org/10.1145/1321440.1321528

[29] Stephen Soderland, Oren Etzioni, Daniel S Weld, Michael Skinner, Jeff Bilmes, and others. 2009. Compiling a massive, multilingual dictionary via probabilistic inference. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Volume 1*. Association for Computational Linguistics, 262–270.

[30] Morris Swadesh. 1955. Towards greater accuracy in lexicostatistic dating. *Int. J. Am. Ling.* 21, 2 (1955), 121–137.

[31] Kumiko Tanaka and Kyoji Umemura. 1994. Construction of a bilingual dictionary intermediated by a third language. In *Proceedings of the 15th Conference on Computational Linguistics, Volume 1*. Association for Computational Linguistics, 297–303.

[32] Rie Tanaka, Yohei Murakami, and Toru Ishida. 2009. Context-based approach for pivot translation services. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'09)*, Vol. 2009. 1555–1561.

[33] Jörg Tiedemann. 2009. Character-based PSMT for closely related languages. In *Proceedings of the 13th Conference of the European Association for Machine Translation (EAMT'09)*. 12–19.

[34] Renee Van Bezooijen and Charlotte Gooskens. 2005. How easy is it for speakers of dutch to understand frisian and afrikaans, and why? *Ling. Netherlands* 22, 1 (2005), 13–24.

[35] Hua Wu and Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. *Mach. Transl.* 21, 3 (01 Sep 2007), 165–181. DOI: http://dx.doi.org/10.1007/s10590-008-9041-6

[36] Mairidan Wushouer, Donghui Lin, Toru Ishida, and Katsutoshi Hirayama. 2014. *Pivot-Based Bilingual Dictionary Extraction from Multiple Dictionary Resources*. Springer International, Cham, 221–234. DOI: http://dx.doi.org/10.1007/978-3-319-13560-1_18

[37] Mairidan Wushouer, Donghui Lin, Toru Ishida, and Katsutoshi Hirayama. 2015. A constraint approach to pivot-based bilingual dictionary induction. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 15, 1, Article 4 (Nov. 2015), 26 pages. DOI: http://dx.doi.org/10.1145/2723144

Received February 2017; revised August 2017; accepted September 2017