Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

The CQC Algorithm: Cycling in Graphs to Semantically Enrich and Enhance a Bilingual Dictionary (Extended Abstract)*

Tiziano Flati and **Roberto Navigli**
Dipartimento di Informatica
Sapienza Università di Roma
{flati, navigli}@di.uniroma1.it

Abstract

Bilingual machine-readable dictionaries are knowledge resources useful in many automatic tasks. However, compared to monolingual computational lexicons like WordNet, bilingual dictionaries typically provide a lower amount of structured information such as lexical and semantic relations, and often do not cover the entire range of possible translations for a word of interest. In this paper we present Cycles and Quasi-Cycles (CQC), a novel algorithm for the automated disambiguation of ambiguous translations in the lexical entries of a bilingual machine-readable dictionary.

1 Introduction

Lexical knowledge resources, such as thesauri, machine-readable dictionaries, computational lexicons and encyclopedias, have been enjoying increasing popularity over the last few years. Among such resources we cite Roget's Thesaurus, the Macquarie Thesaurus, the Longman Dictionary of Contemporary English [Proctor, 1978, LDOCE], WordNet [Fellbaum, 1998] and Wikipedia. These knowledge resources have been utilized in many applications, including Word Sense Disambiguation (WSD) [Yarowsky, 1992; Nastase and Szpakowicz, 2001; Martínez *et al.*, 2008, cf. Navigli, 2009b, 2012 for a survey], Semantic Information Retrieval [Krovetz and Croft, 1992; Mandala *et al.*, 1998; Sanderson, 2000, inter alia], Question Answering [Lita *et al.*, 2004] and knowledge acquisition [Navigli and Ponzetto, 2012].

Most of these applications exploit the structure provided by the adopted lexical resources in a number of different ways. For instance, lexical and semantic relations encoded in computational lexicons such as WordNet have been shown to be very useful in graph-based WSD [Mihalcea, 2005; Agirre and Soroa, 2009; Navigli and Lapata, 2010; Ponzetto and Navigli, 2010] and semantic similarity [Pedersen *et al.*, 2005]. Interestingly, it has been reported that the higher the amount of structured knowledge, the higher the disambiguation performance [Navigli and Lapata, 2010].

*This paper is an extended abstract of the JAIR publication [Flati and Navigli, 2012].

Unfortunately, not all the semantics are made explicit within lexical resources. Even WordNet, the most widely-used computational lexicon of English, provides explanatory information in the unstructured form of textual definitions, i.e., strings of text which explain the meaning of concepts using possibly ambiguous words (e.g., "motor vehicle with four wheels" is provided as a definition of the most common sense of *car*). Still worse, while computational lexicons like WordNet contain semantically explicit information such as *is-a* and *part-of* relations, machine-readable dictionaries (MRDs) are often just electronic transcriptions of their paper counterparts. Thus, for each entry they mostly provide implicit information in the form of free text, which cannot be immediately utilized in Natural Language Processing applications. Over recent years various approaches to the disambiguation of monolingual dictionary definitions have been investigated [Harabagiu *et al.*, 1999; Litkowski, 2004; Castillo *et al.*, 2004; Navigli and Velardi, 2005; Navigli, 2009a], and results have shown that they can, indeed, boost the performance of difficult tasks such as WSD [Cuadros and Rigau, 2008; Agirre and Soroa, 2009]. However, little attention has been paid to the disambiguation of bilingual dictionaries, which would be capable of improving popular applications such as Machine Translation.

In this paper we present a graph-based algorithm which aims at disambiguating translations in bilingual machine-readable dictionaries. Taken as input a bilingual MRD, our method transforms the dictionary into a graph whose nodes are word senses and edges encode translation relations. Next, we introduce a novel notion of cyclic and quasi-cyclic graph paths that we use to select the most appropriate sense for a translation w' of a source word w.

2 Preliminaries

Goal. The general form of a bilingual dictionary entry is:

$$w_p^i \rightarrow v_1, v_2, \ldots, v_k$$

where: (i) w_p^i is the i-th sense of the word w with part of speech p in the source language (e.g., $play_v^2$ is the second sense of the verb *play*); (ii) each v_j is a translation in the target language for sense w_p^i (e.g., $suonare_v$ is a translation

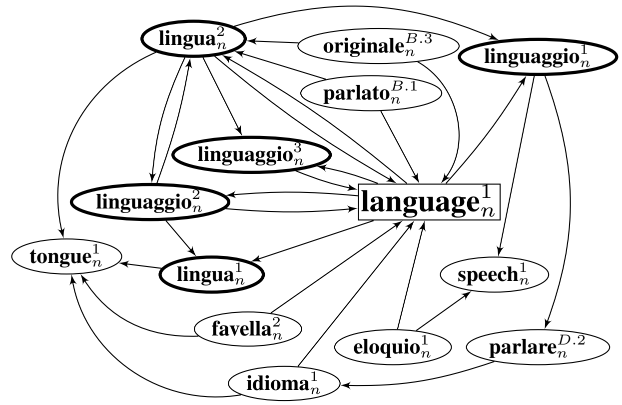

Figure 1: An excerpt of the Ragazzini-Biagi noisy graph including *language*^1_n and its neighbours.

for *play*^2_v). Note that each v_j is implicitly assumed to have the same part of speech p as w_p. Importantly, no sense is explicitly associated with v_j.

Our objective is to associate, in a systematic and automatic way, each target word v_j with one of its senses so that the concepts expressed by w_p and v_j match.

Bilingual Dictionary. We define a **bilingual machine-readable dictionary** (BiMRD) as a quadruple $D = (\mathcal{L}, Senses, \mathcal{T}, \mathcal{M})$, where $\mathcal{L}$ is the bilingual lexicon (i.e., $\mathcal{L}$ includes all the lexical items for both languages), $Senses$ is a mapping such that, given a lexical item $w \in \mathcal{L}$, returns the set of senses for w in D, $\mathcal{T}$ is a translation function which, given a word sense $s \in Senses(w)$, provides a set of (possibly ambiguous) translations for s. Typically, $\mathcal{T}(s) \subset \mathcal{L}$, that is, the translations are in the lexicon. However, it might well be that some translations in $\mathcal{T}(s)$ are not in the lexicon. Finally, $\mathcal{M}$ is a function which, given a word sense $s \in Senses(w)$, provides the set of all words representing meta-information for sense s (e.g., $\mathcal{M}$(*phoneme*^1_n) = {*linguistics*}).

The dictionary also provides usage examples and compound translations , lexical variants (e.g., *acknowledgement* vs. *acknowledgment*) and references to other entries (e.g., from *motorcar* to *car*).

Noisy Graph. Given a BiMRD D, we define a **noisy dictionary graph** $G = (V, E)$ as a directed graph where:

1. V is the set of senses in the dictionary D (i.e., $V = \bigcup_{w \in \mathcal{L}} Senses(w)$);
2. For each word $w \in \mathcal{L}$ and a sense $s \in Senses(w)$, an edge (s, s') is in E if and only if s' is a sense of a translation of s in the dictionary (i.e., $s' \in Senses(w')$ and $w' \in \mathcal{T}(s)$), or s' is a sense of a meta-word m in the definition of s (i.e., if $s' \in Senses(m)$ for some $m \in \mathcal{M}(s)$).

Graph Cycles and Quasi-Cycles. We now recall the definition of **graph cycle**. A cycle for a graph G is a sequence of edges of G that form a path $v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_n$ ($v_i \in V\ \forall i \in \{1, \ldots, n\}$) such that the first node of the path corresponds to the last, i.e., $v_1 = v_n$. The length of a cycle is given by the number of its edges.

We further provide the definition of **quasi-cycle** as a sequence of edges in which the reversal of the orientation of one or more consecutive edges creates a cycle [Bohman and Thoma, 2000].

It can be seen that the reversal of an edge of a quasi-cycle creates a cycle. Since the direction of this edge is opposite to that of the cycle, we call it a reversed edge. Finally, we say that a path is (quasi-)cyclic if it forms a (quasi-)cycle. Note that we do not consider paths going across senses of the same word.

3 The CQC Algorithm

We are now ready to introduce the Cycles & Quasi-Cycles (CQC) algorithm, whose pseudocode is given in Table 1. The algorithm takes as input a BiMRD $D = (\mathcal{L}, Senses, \mathcal{T}, \mathcal{M})$, and a sense s of a word w in its lexicon (i.e., $w \in \mathcal{L}$ and $s \in Senses(w)$). The algorithm aims at disambiguating each of the word's ambiguous translations $w' \in \mathcal{T}(s)$, i.e., to assign it the right sense among those listed in $Senses(w')$.

The algorithm outputs a mapping μ between each ambiguous word $w' \in \mathcal{T}(s)$ and the sense s' of w' chosen as a result of the disambiguation procedure that we illustrate hereafter.

First, for each sense s' of our target translation $w' \in \mathcal{T}(s)$, the algorithm performs a search of the noisy graph associated with D and collects the following kinds of paths:

i) Cycles: $s \rightarrow s' \rightarrow s_1 \rightarrow \cdots \rightarrow s_{n-2} \rightarrow s_{n-1} = s$

ii) Quasi-cycles:

$$s \rightarrow s' \rightarrow s_1 \rightarrow ... \rightarrow s_j \leftarrow ... \leftarrow s_k \rightarrow ... \rightarrow s_{n-1} = s \quad (1)$$

$$1 \leq j \leq n-2, j < k \leq n-1$$

where s is our source sense, s' is our candidate sense for $w' \in \mathcal{T}(s)$, s_i is a sense listed in D ($i \in \{1, \ldots, n-2\}$), $s_{n-1} = s$, and n is the length of the path. Note that both kinds of path start and end with the same node s, and that the algorithm searches for quasi-cycles whose reversed edges connecting s_k to s_j are consecutive. To avoid redundancy we require (quasi-)cycles to be simple, that is, no node is repeated in the path except the start/end node (i.e., $s_i \neq s, s_i \neq s', s_i \neq s_{i'}\ \forall i, i'$ s. t. $i \neq i'$).

The second phase of the CQC algorithm (lines 5-10 of Table 1) computes a score for each sense s' of w' based on the paths collected for s' during the first phase. Let p be such a path, and let l be its length. Then the contribution of p to the score of s' is given by:

$$score(p) := \frac{\omega(l)}{NumPaths(all\_paths, l)} \quad (2)$$

where: (i) $\omega(l)$ is a monotonically non-increasing function of its length l; in our experiments, we tested three different weight functions $\omega(l)$, namely a constant, a linear and an inversely exponential function[1]; (ii) the normalization fac-

[1]Experiments showed the inversely exponential function to be the best performing one.

Dataset	# entries	# translations	# polysemous	avg. polysemy	perfect alignments
TUNING DATASET	50	80	53	4.74	37
TEST DATASET	500	1,069	765	3.95	739

Table 2: Statistics for the tuning and test datasets.

```
CQC(BiMRD D = (L, Senses, T, M), sense s of w ∈ L)
for each word w' ∈ T(s)
    for each sense s' ∈ Senses(w')
        paths(s') := DFS(s', s)
    all_paths := ⋃_{s' ∈ Senses(w')} paths(s')
    for each sense s' ∈ Senses(w')
        score(s') := 0
        for each path p ∈ paths(s')
            l := length(p)
            v := ω(l) · 1 / NumPaths(all_paths, l)
            score(s') := score(s') + v
    μ(w') = argmax_{s' ∈ Senses(w')} score(s')
return μ
```

Table 1: The Cycles & Quasi-Cycles (CQC) algorithm in pseudocode.

tor $NumPaths(all\_paths, l)$ calculates the overall number of collected paths of length l among all the target senses.

As a result of the systematic application of the algorithm to each sense in our BiMRD D, a new graph $G' = (V, E')$ is output, where V is again the sense inventory of D, and E' is a subset of the noisy edge set E such that each edge $(s, s') \in E'$ is the result of our disambiguation algorithm run with input D and s. In this pruned graph, each sense links to only one sense of each of its translations.

4 Evaluation: Dictionary Disambiguation

Dictionary. We performed our dictionary disambiguation experiments on the Ragazzini-Biagi [Ragazzini and Biagi, 2006], a popular bilingual English-Italian dictionary, which contains over 90,000 lemmas and 150,000 word senses.

Dataset. Our datasets for tuning and test consist of dictionary entries, each containing translations of a source sense into a target language. Each translation item was manually disambiguated according to its sense inventory in the bilingual dictionary.

We report statistics for the two datasets in Table 2, including the number of polysemous translations and the average polysemy of each translation. We note that for 44 of the translations in the test set (i.e., 4.1% of the total) none of the senses listed in the dictionary is appropriate (including monosemous translations). The last column in the table shows the number of translations for which a sense exists that translates back to the source lemma.

Algorithms. We compared the following algorithms in our experimental framework, since they represent the most widespread graph-based approaches and are used in many NLP tasks with state-of-the-art performance:

- **CQC**: we applied the CQC algorithm as described in Section 3.
- **Cycles**, a variant of the CQC algorithm which searches for cycles only (i.e., quasi-cycles are not collected).
- **DFS**, which applies an ordinary DFS algorithm and collects all paths between s and s'.
- **Random walks**, which performs a large number of random walks starting from s', collecting those paths that lead to s.
- **Markov chains**, which calculates the probability of arriving at a certain source sense s starting from the initial translation sense s' averaged over n consecutive steps. The initial Markov chain is initialized so that the outgoing edges of a node have equal probabilities.
- **Personalized PageRank (PPR)**: a popular variant of the PageRank algorithm [Brin and Page, 1998] where the original Markov chain approach to node ranking is modified by perturbating the initial probability distribution on nodes. We concentrate all the probability mass on s', apply PPR and select the best translation sense as the one which maximizes the PPR value of the source word.
- **Lesk algorithm** [Lesk, 1986]: we apply an adaptation of the Lesk algorithm in which, given a source sense s of word w and a word w' occurring as a translation of s, we determine the right sense of w' on the basis of the (normalized) maximum overlap between the entries of each sense s' of w' and that of s.

We also compared the performance of our algorithms with three baselines: (i) the **First Sense (FS) Baseline**, that associates the first sense listed by the dictionary with each translation to be disambiguated; (ii) the **Random Baseline**, which selects a random sense for each target translation; (iii) the **Degree Baseline**, that chooses the translation sense with the highest out-degree.

Parameters. We used the tuning dataset to fix the parameters of each algorithm that maximized the performance (see Flati and Navigli, 2012 for details).

Measures. To assess the performance of our algorithms, we calculated precision (the number of correct answers over the number of items disambiguated by the system), recall (the number of correct answers over the number of items in the dataset), and F1 (a harmonic mean of precision and recall, given by $\frac{2PR}{P+R}$). Note that precision and recall do not consider those items in the test set for which no appropriate sense is available in the dictionary. In order to account for these items, we also calculated accuracy as the number of correct answers divided by the total number of items in the test set.

Algorithm	P	R	F1	A
CQC	87.14	83.32	85.19	83.35
Cycles	87.17	74.93	80.59	75.58
DFS	63.40	37.85	47.40	39.85
Random walks	83.94	61.17	70.77	62.49
Markov chains	85.46	65.37	74.08	66.70
PPR	83.20	81.25	82.21	81.27
Lesk	86.05	31.90	46.55	34.52
First Sense BL	72.67	73.17	72.92	73.53
Random BL	28.53	29.76	29.13	28.53
Degree BL	58.39	58.85	58.39	58.62

Table 3: Disambiguation performance on the Ragazzini-Biagi dataset.

Results. In Table 3 we report the results of our algorithms on the test set. CQC, PPR and Cycles are the best performing algorithms, achieving around 83%, 81% and 75% accuracy respectively. CQC outperforms all other systems in terms of F1 by a large margin. The results show that the mere use of cyclic patterns does not lead to state-of-the-art performance, which is obtained when quasi-cycles are also considered. Including quasi-cycles leads to a considerable increase in recall, while maintaining a high level of precision. The DFS is even more penalizing because it does not get backward support as happens for cycling patterns. Markov chains consistently outperform Random walks. We hypothesize that this is due to the higher coverage of Markov chains compared to the number of random walks collected by a simulated approach. PPR outperforms the two other probabilistic approaches, but lags behind CQC by 3 points in F1 and 2 in accuracy. This result confirms previous findings concerning the high performance of PPR, but also corroborates our hunch about quasi-cycles being the determining factor in the detection of semantic connections. Finally, Lesk achieves high precision, but low recall, due to the lack of a lookahead mechanism.

The random baseline represents our lowerbound and is much lower than all other results. Compared to the first sense baseline, CQC, PPR and Cycles obtain better performance. This result is consistent with previous findings for tasks such as the Senseval-3 Gloss Word Sense Disambiguation [Litkowski, 2004]. However, at the same time, it is in contrast with results on all-words WSD [Navigli, 2009b], where the first or most frequent sense baseline generally outperforms most disambiguation systems. Nevertheless, the nature of these two tasks is very different, because in dictionary entries senses tend to be equally distributed, whereas in open text they have a single predominant meaning that is determined by context. As for the Degree Baseline, it yields results below expectations, and far worse than the FS baseline. The reason behind this lies in the fact that the amount of translations and translation senses does not necessarily correlate with mainstream meanings.

5 Related Work

Since the late 1970s much work on the analysis and disambiguation of dictionary glosses has been done. This includes methods for the automatic extraction of taxonomies from lexical resources [Litkowski, 1978; Amsler, 1980], the identification of genus terms [Chodorow *et al.*, 1985] and the extraction of explicit information from machine-readable dictionaries ([Nakamura and Nagao, 1988; Ide and Véronis, 1993], as well as the construction of ambiguous semantic networks from glosses [Kozima and Furugori, 1993].

More recently, a set of heuristics has been proposed to semantically annotate WordNet glosses, leading to the release of the eXtended WordNet [Harabagiu *et al.*, 1999; Moldovan and Novischi, 2004]. Among the heuristics, the cross reference heuristic is the closest technique to our notion of (quasi-)cyclic patterns in which cycles of length 2 are sought. Recently, it has been proposed that probabilistic translation circuits can be used to automatically acquire a multilingual dictionary [Mausam *et al.*, 2009].

Based on the eXtended WordNet, a gloss disambiguation task was organized at Senseval-3 [Litkowski, 2004]. Most notably, the best performing systems, namely the TALP system [Castillo *et al.*, 2004], and SSI [Navigli and Velardi, 2005], are knowledge-based and rely on rich knowledge resources: respectively, the Multilingual Central Repository [Atserias *et al.*, 2004], and a proprietary lexical knowledge base (cf. Navigli and Lapata, 2010).

However, the literature in the field of gloss disambiguation is focused only on monolingual dictionaries, such as WordNet and LDOCE, while, to our knowledge, CQC is the first algorithm aimed at disambiguating the entries of a bilingual dictionary. Moreover, in contrast to many disambiguation methods in the literature [Navigli, 2009b], our approach does not exploit lexical and semantic relations, such as those available in computational lexicons like WordNet.

6 Conclusions

In this paper we presented a novel algorithm, called Cycles and Quasi-Cycles (CQC), for the disambiguation of bilingual machine-readable dictionaries. The algorithm is based on the identification of (quasi-)cycles in the noisy dictionary graph, i.e., circular edge sequences (possibly with some consecutive edges reversed) relating a source word sense to a target one.

We show that our notion of (quasi-)cyclic patterns enables state-of-the-art performance to be attained in the disambiguation of dictionary entries, surpassing all other disambiguation approaches, as well as a competitive baseline such as the first sense heuristic. Crucially, the introduction of reversed edges allows us to find more semantic connections, substantially increasing recall while keeping precision very high. The strength of our approach lies in its weakly supervised nature: the CQC algorithm relies exclusively on the structure of the input bilingual dictionary and, unlike others, no further resource is required.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Starting Grant MultiJEDI No. 259234.
The authors wish to thank Zanichelli for making the Ragazzini-Biagi dictionary available for research purposes.

References

[Agirre and Soroa, 2009] Eneko Agirre and Aitor Soroa. Personalizing PageRank for Word Sense Disambiguation. In *Proc. of EACL 2009*, pages 33–41, Athens, Greece, 2009.

[Amsler, 1980] Robert A. Amsler. *The structure of the Merriam-Webster pocket dictionary, Ph.D. Thesis*. University of Texas, Austin, TX, USA, 1980.

[Atserias *et al.*, 2004] Jordi Atserias, Luís Villarejo, German Rigau, Eneko Agirre, John Carroll, Bernardo Magnini, and Piek Vossen. The MEANING Multilingual Central Repository. In *Proc. of GWC 2004*, pages 23–30, Brno, Czech Republic, 2004.

[Bohman and Thoma, 2000] Tom Bohman and Lubos Thoma. A note on sparse random graphs and cover graphs. *The Electronic Journal of Combinatorics*, 7(1):1–9, 2000.

[Brin and Page, 1998] Sergey Brin and Michael Page. Anatomy of a large-scale hypertextual web search engine. In *Proc. of WWW 1998*, pages 107–117, Brisbane, Australia, 1998.

[Castillo *et al.*, 2004] Mauro Castillo, Francis Real, Jordi Asterias, and German Rigau. The TALP systems for disambiguating WordNet glosses. In *Proc. of ACL 2004 SENSEVAL-3 Workshop*, pages 93–96, Barcelona, Spain, 2004.

[Chodorow *et al.*, 1985] Martin Chodorow, Roy Byrd, and George Heidorn. Extracting semantic hierarchies from a large on-line dictionary. In *Proc. of ACL 1985*, pages 299–304, Chicago, IL, USA, 1985.

[Cuadros and Rigau, 2008] Montse Cuadros and German Rigau. KnowNet: Building a large net of knowledge from the web. In *Proc. of COLING 2008*, pages 161–168, Manchester, UK, 2008.

[Fellbaum, 1998] Christiane Fellbaum, editor. *WordNet: An Electronic Database*. MIT Press, Cambridge, MA, 1998.

[Flati and Navigli, 2012] Tiziano Flati and Roberto Navigli. The CQC algorithm: Cycling in graphs to semantically enrich and enhance a bilingual dictionary. *Journal of Artificial Intelligence Research (JAIR)*, 43:135–171, 2012.

[Harabagiu *et al.*, 1999] Sanda Harabagiu, George Miller, and Dan Moldovan. WordNet 2 - a morphologically and semantically enhanced resource. In *Proc. of SIGLEX 1999*, pages 1–8, Maryland, USA, 1999.

[Ide and Véronis, 1993] Nancy Ide and Jean Véronis. Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time? In *Proc. of Workshop on Knowledge Bases and Knowledge Structures*, pages 257–266, Tokyo, Japan, 1993.

[Kozima and Furugori, 1993] Hideki Kozima and Teiji Furugori. Similarity between words computed by spreading activation on an english dictionary. In *Proc. of ACL 1993*, pages 232–239, Utrecht, The Netherlands, 1993.

[Krovetz and Croft, 1992] Robert Krovetz and William B. Croft. Lexical ambiguity and Information Retrieval. *ACM Transactions on Information Systems*, 10(2):115–141, 1992.

[Lesk, 1986] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proc. of SIGDOC 1986*, pages 24–26, New York, NY, 1986.

[Lita *et al.*, 2004] Lucian Vlad Lita, Warren A. Hunt, and Eric Nyberg. Resource analysis for Question Answering. In *Proc. of ACL 2004, Interactive poster and demonstration sessions*, pages 162–165, Barcelona, Spain, 2004.

[Litkowski, 1978] Kenneth C. Litkowski. Models of the semantic structure of dictionaries. *American Journal of Computational Linguistics*, 81:25–74, 1978.

[Litkowski, 2004] Kenneth C. Litkowski. SENSEVAL-3 task: Word-Sense disambiguation of WordNet glosses. In *Proc. of ACL 2004 SENSEVAL-3 Workshop*, pages 13–16, Barcelona, Spain, 2004.

[Mandala *et al.*, 1998] Rila Mandala, Takenobu Tokunaga, and Hozumi Tanaka. The use of WordNet in Information Retrieval. In *Proc. of COLING/ACL Workshop on Usage of WordNet in NLP Systems*, pages 31–37, Montreal, Canada, 1998.

[Martínez *et al.*, 2008] David Martínez, Oier Lopez de Lacalle, and Eneko Agirre. On the use of automatically acquired examples for all-nouns Word Sense Disambiguation. *Journal of Artificial Intelligence Research (JAIR)*, 33:79–107, 2008.

[Mausam *et al.*, 2009] Mausam, Stephen Soderland, Oren Etzioni, Daniel Weld, Michael Skinner, and Jeff Bilmes. Compiling a massive, multilingual dictionary via probabilistic inference. In *Proc. of ACL-IJCNLP 2009*, pages 262–270, Singapore, 2009.

[Mihalcea, 2005] Rada Mihalcea. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proc. of HLT/EMNLP*, pages 411–418, Vancouver, BC, 2005.

[Moldovan and Novischi, 2004] Dan Moldovan and Adrian Novischi. Word Sense Disambiguation of WordNet glosses. *Computer Speech & Language*, 18(3):301–317, 2004.

[Nakamura and Nagao, 1988] Jun-Ichi Nakamura and Makoto Nagao. Extraction of semantic information from an ordinary English dictionary and its evaluation. In *Proc. of COLING 1988*, pages 459–464, Budapest, Hungary, 1988.

[Nastase and Szpakowicz, 2001] Vivi Nastase and Stan Szpakowicz. Word Sense Disambiguation in Roget's Thesaurus Using WordNet. In *Proc. of the NAACL WordNet and Other Lexical Resources workshop*, pages 17–22, Pittsburgh, USA, 2001.

[Navigli and Lapata, 2010] Roberto Navigli and Mirella Lapata. An experimental study on graph connectivity for unsupervised Word Sense Disambiguation. *IEEE TPAMI*, 32(4):678–692, 2010.

[Navigli and Ponzetto, 2012] Roberto Navigli and Simone Paolo Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.

[Navigli and Velardi, 2005] Roberto Navigli and Paola Velardi. Structural semantic interconnections: a knowledge-based approach to Word Sense Disambiguation. *IEEE TPAMI*, 27(7):1075–1088, 2005.

[Navigli, 2009a] Roberto Navigli. Using Cycles and Quasi-Cycles to disambiguate dictionary glosses. In *Proc. of EACL 2009*, pages 594–602, 2009.

[Navigli, 2009b] Roberto Navigli. Word Sense Disambiguation: a Survey. *ACM Computing Surveys*, 41(2):1–69, 2009.

[Navigli, 2012] Roberto Navigli. A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches. In *Proc. of SOFSEM 2012*, Špindlerův Mlýn, Czech Republic, 2012.

[Pedersen *et al.*, 2005] Ted Pedersen, Satanjeev Banerjee, and Siddharth Patwardhan. Maximizing semantic relatedness to perform Word Sense Disambiguation. In *UMSI Research Report 2005/25*, Minnesota, 2005.

[Ponzetto and Navigli, 2010] Simone Paolo Ponzetto and Roberto Navigli. Knowledge-rich Word Sense Disambiguation rivaling supervised system. In *Proc. of ACL 2010*, pages 1522–1531, Uppsala, Sweden, 2010.

[Proctor, 1978] Paul Proctor, editor. *Longman Dictionary of Contemporary English*. Longman Group, UK, 1978.

[Ragazzini and Biagi, 2006] Giuseppe Ragazzini and Adele Biagi, editors. *Il Ragazzini-Biagi,* 4^{th} *Edition*. Zanichelli, Italy, 2006.

[Sanderson, 2000] M. Sanderson. Retrieving with good sense. *Information Retrieval*, 2(1):49–69, 2000.

[Yarowsky, 1992] David Yarowsky. Word-sense disambiguation using statistical models of the Roget's categories trained on large corpora. In *Proc. of COLING 1992*, Nantes, France, 1992.