Research Statement

Mohammad Taher Pilehvar
Language Technology Lab,
University of Cambridge.
mp792@cam.ac.uk

This statement outlines my current and future research directions and provides a brief summary of my past research, with the hope that it elucidates my research contributions and academic ambitions.

Introduction

I am currently a research associate at the Language Technology Lab in the University of Cambridge. I work with Dr. Nigel Collier on PheneBank, a project funded by the UK's Medical Research Council which seeks to apply Natural Language Processing (NLP) techniques, such as named entity recognition, text classification, and harmonization, to medical texts. Prior to this, I did research in Dr. Roberto Navigli's LCL lab in the Sapienza University of Rome within MultiJEDI, an ERC funded project.

The main theme of my research is semantic representation of word senses and concepts. The objective of this research is to address one of the main limitations of word representations, i.e., meaning conflation deficiency (see Section 3), through modelling individual meanings (senses) of words. These representations can be integrated into downstream NLP systems in order to resolve lexical ambiguity in their input which can ease their challenge in natural language understanding.

My research interests mainly lie within NLP, particularly Lexical Semantics. Broadly speaking, I have contributed to the following areas of research in NLP:

- **Semantic representation** (§3)
  - Word senses and concepts
  - Rare and unseen words
  - Multilingual
- **Semantic similarity measurement** (§4)
  - Sense level similarity
  - Alignment-based similarity of words, phrases and sentences
  - Cross-level, cross-lingual and multi-lingual
- **Word Sense Disambiguation (WSD)** (§5)
  - Polysemy simulation
  - Large-scale evaluation of WSD
- **Ontology enrichment and alignment** (§6)
  - Ontologization of collaborative resources
  - Alignment of heterogeneous lexical resources
- **Statistical Machine Translation** (§7)
  - Automatic construction of parallel corpora
  - English-Persian phrase-based machine translation

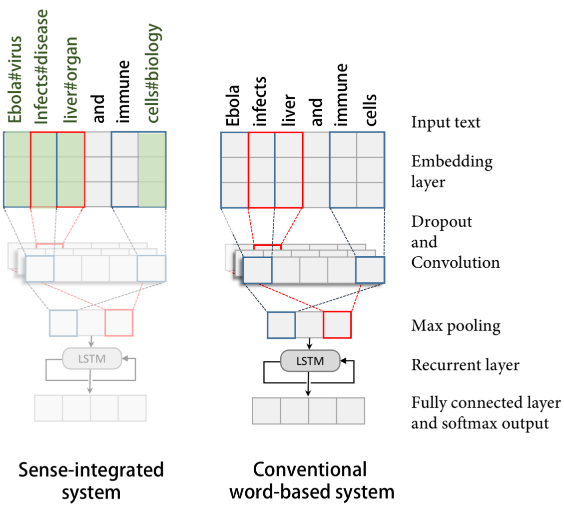

Figure 1
A simplified illustration of a sense-integrated NLP system (a CNN-LSTM neural classifier). The integration mainly happens on the embedding layer and through resolving input ambiguity.

- **Geographical parsing** (§8)
  - Metonymy detection
  - Geo-coding

Ongoing and Future Research

1. Sense-aware NLP

As a general trend, most current NLP systems function at the word level, i.e. individual words constitute the most fine-grained meaning-bearing elements of their input. The word level functionality can affect the performance of these systems in two ways: (1) it can hamper their efficiency in handling words that are not encountered frequently during training, such as multiwords, inflections and derivations, and (2) it can restrict their semantic understanding to the level of words, with all their ambiguities, and thereby prevent accurate capture of the intended meanings.

The first issue can be alleviated by inducing embeddings for rare and unseen words, for instance, by exploiting the knowledge encoded in external lexical resources (see Section 3.2). A research direction of mine focuses on the second issue: making NLP systems sense-aware. Despite its potential benefits, the integration of sense-level information into NLP systems has remained largely understudied, perhaps for not being straightforward process. Downstream NLP applications often take as their input a sequence of words, ignoring the fact that lexical ambiguity can hinder accurate semantic understanding or just hoping that the issue can be automatically addressed with the abundance of data. However, often this cannot be achieved by the model given the complexity of decisions or scarcity of training data. Examples of non-optimal sense

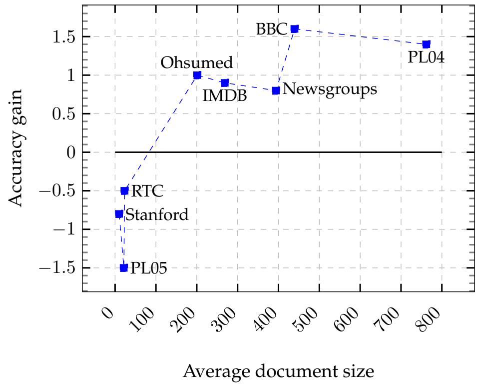

Figure 2
Relation between average document size and performance improvement in sense-integrated classifier. Results show that sense integration leads to consistent improvements as document size grows.

distinction can be found across a wide range of NLP applications, including celebrated commercial products such Google Translate[1].

As a very first step towards this direction, we carried out a set of experiments showing that a simple disambiguation coupled with the integration of sense representations can lead to consistent performance improvement on multiple topic categorization and polarity detection datasets, particularly when the fine granularity of the underlying sense inventory was reduced and the document was sufficiently large (Figure 2). The results of this study were presented at ACL 2017 [1]. Given these encouraging results and in light of the potential advantages of functioning at the deep semantic level of senses, a future research direction of mine will investigate fusing sense level information into various downstream applications, particularly machine translation systems and chatbots which can potentially benefit significantly from this migration.

2. Semantic Space Unification

One of the keys to word embeddings' recent success is their dense vectorial representation. This simple numerical representation allows their seamless integration into NLP systems, particularly in the input layer of various neural architectures. Replacing one-hot representation of words in this layer with soft continuous representations of embeddings can provide the system with enhanced generalization power and lead to significant performance improvements. However, wide-coverage lexical semantic knowledge is not bound to the distributional form derived from co-occurrence statistics in text corpora. Importantly, there exist hundreds of lexical resources, such as machine-readable dictionaries, ontologies, thesauri, and semantic networks, for various domains and languages. Usually created by experts or through a collaborative endeavour, these lexical resources provide a wealth of knowledge which might not be easy to capture

1 For instance, in the EN-IT translation, the word *plant* in "he works in a *plant*" is translated to *pianta* which refers to the botanical plant, or the term *showers* in "many showers expected according to the forecast" is translated to *docce* which refers to the bathroom shower.

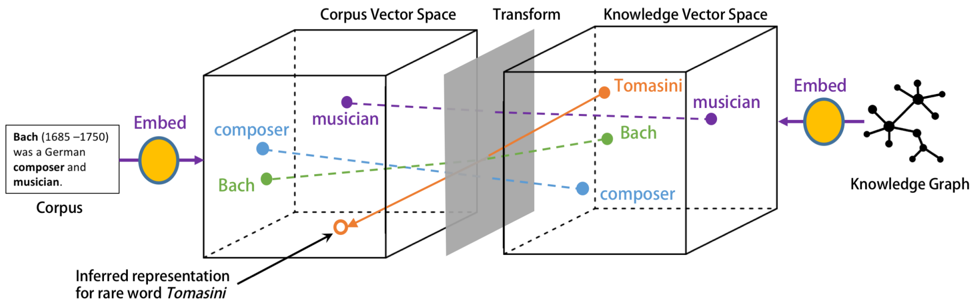

Figure 3
Merging heterogeneous semantic spaces (e.g., corpus-based and graph-based) with dual objectives: knowledge base completion (not shown in the figure) and rare word representation (for the word "Tomasini" in the figure).

using distributional modeling. Hence, these resources can be seen as complementary sources of lexical knowledge to distributional models.

A possible future research direction of mine would focus on the unification of these two types of knowledge representations with the goal of improving individual spaces. Essentially, this problem can be viewed as two sub-problems, described in the following.

2.1 Embedding of lexical resources

Knowledge in lexical resources is usually encoded either in terms of binary lexico-semantic relations, or through structures of hierarchies or semantic networks. This non-numerical representation usually impedes their easy integration into NLP systems, particularly given the heterogeneity of lexical resources and the type of knowledge they provide. One way to address this issue is to transform (embed) knowledge encoded in these resources into a semantic space. The embedding procedure can also be guided by the target task at hand and the type of lexical information which is most important for the task. The works done in [2, 3] are in this path. However, there remains plenty of room for improving these techniques, particularly in capturing different type of knowledge encoded in a single model and in densifying the resulting space.

2.2 Alignment of heterogeneous semantic spaces

Having a lexical resource represented in the form of a vectorial semantic space eases the process of aligning it with distributional models. A current research focus of mine is on the alignment of heterogeneous semantic spaces which are either constructed from different sources or exhibit different properties. The alignment can be advantageous to several NLP applications. For instance, in [4] we showed that a linear alignment of corpus-based and knowledge base spaces can be used to induce new embeddings for rare and unseen words in the former space (Figure 3 shows an illustration). Viewing the alignment in the inverse direction, the procedure can be leveraged for knowledge base completion.

Past Research

3. Semantic Representation

Semantic representation, i.e., modeling the meanings (semantics) of linguistic items in a mathematical or machine interpretable form, is a fundamental problem and a key research challenge in NLP and Artificial Intelligence. Currently, the most prominent representation approach is the Vector Space Model (VSM) which views a linguistic item as a vector (or a point) in an n-dimensional semantic space. The semantic similarity between points is then defined in terms of their distances (such as Euclidean, cosine, etc.). Research in semantic representation has recently experienced a resurgence of interest with neural network-based models that view the representation task as a language modeling problem and learn dense representations (usually referred to as embeddings) by efficiently processing massive amounts of texts.

Meaning conflation deficiency. Either in its conventional count-based form or the recent predictive approach, the prevailing objective of representing each word type as a single point in the semantic space has a major limitation (i.e., meaning conflation deficiency): words can have multiple meanings. Representing a potentially polysemous word as a single point in the semantic space conflates all these meanings into a single representation. This objective can have negative impacts on accurate semantic modeling, e.g., semantically unrelated words that are synonymous to different senses of a word are pulled towards each other in the semantic space. For example, the two semantically-unrelated words squirrel and keyboard are pulled towards each other in the semantic space for their similarities to two different senses of mouse, i.e., rodent and computer input device.

3.1 Word senses and concepts

Semantic representation of word senses and concepts comes as a solution to the meaning conflation deficiency of word representations. Because they represent the most fine-grained semantic level of language, word senses play a vital role in natural language understanding. Effective representations of these entities can be directly useful to Word Sense Disambiguation, semantic similarity, coarsening sense inventories, alignment of lexical resources, lexical substitution, and semantic priming. Moreover, sense-level representation can be directly extended to applications requiring word representations, with the added benefit that they provide an extra level of semantic distinction.

3.1.1 Align, Disambiguate, and Walk (ADW). One way to model word senses is to view their sense inventory as a semantic network and directly exploit lexical and structural knowledge in this network for modeling individual word senses. ADW [2] performs a series of random walks on the semantic network of WordNet in order to represent individual nodes (i.e., synsets) in this network. ADW is coupled with an alignment-based disambiguation technique that transforms a given pair of texts to their intended meanings. In [5], we extended ADW to the semantic network of Wiktionary and carried out an extensive evaluation and analysis on several benchmarks.

3.1.2 De-Conflated Semantic Representations (DeConf). Another way to compute sense embeddings is to do a post-processing on word embeddings. By exploiting deep knowledge from the semantic network of WordNet, DeConf [6] breaks a given word

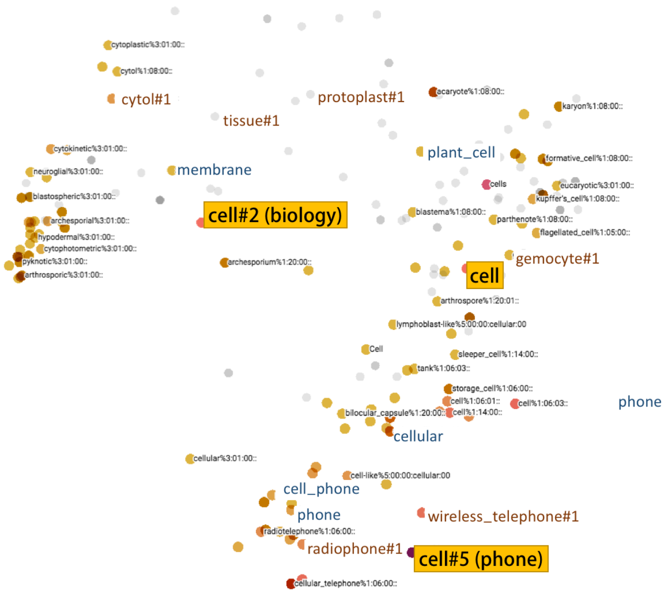

Figure 4
A 2D illustration of a unified semantic space of words and word senses.

representation into its constituent sense representations. DeConf provides multiple advantages in comparison to the past work, mainly for its linkage to a standard sense inventory and its deep exploitation of this resource.

DeConf gets a d-dimensional pre-trained set of word embeddings and computes sense embeddings in the same semantic space (Figure 4 shows an illustration of this space). Let $\mathcal{V}$ represent this set. Our objective here is to compute a set $\mathcal{V}^* = \{v^*_{s_1}, \ldots, v^*_{s_n}\}$ of representations for n word senses $\{s_1, \ldots, s_n\}$ in the same d-dimensional semantic space of words. We achieve this for each sense s_i by de-conflating the representation v_{s_i} of its corresponding lemma and biasing it towards the representations of the words in $\mathcal{B}_i$. Specifically, we obtain a representation $v^*_{s_i}$ for a word sense s_i by solving:

$$\arg\min_{v^*_{s_i}} \alpha \; d(v^*_{s_i}, v_{s_i}) + \sum_{b_{ij} \in \mathcal{B}_i} \delta_{ij} \; d(v^*_{s_i}, v_{b_{ij}}) \qquad (1)$$

where v_{s_i} and $v_{b_{ij}}$ are the respective word representations ($\in \mathcal{V}$) of the lemma of s_i and the j^{th} biasing word in the list of biasing words for s_i, i.e, $\mathcal{B}_i$. The distance $d(v, v')$ between vectors v and v' is measured by squared Euclidean distance $\|v - v'\|^2 = \sum_k (v_k - v'_k)^2$. The first term in Formula 1 requires the representation of the word sense s_i (i.e., $v^*_{s_i}$) to be similar to that of its corresponding lemma, i.e., v_{s_i}, whereas the second term encourages $v^*_{s_i}$ to be in the proximity of its biasing words in the semantic space. The above criterion can be solved in an iterative manner and by computing the

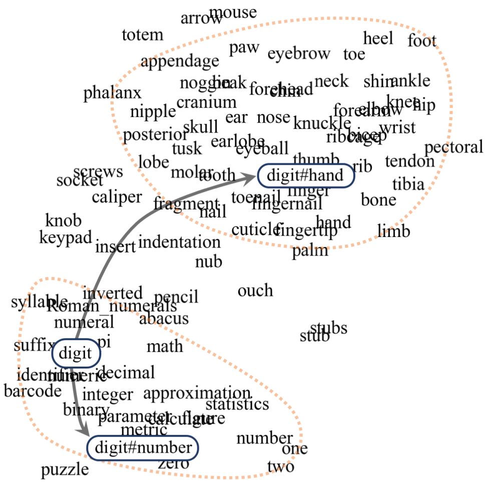

Figure 5
The illustration of the word *digit* and two of its computed senses in our unified 2D semantic space.

representation of a word sense s_i:

$$v^*_{s_i} = \frac{\alpha v_{s_i} + \sum_{b_{ij} \in \mathcal{B}_i} \delta_{ij} v_{b_{ij}}}{\alpha + \sum_j \delta_{ij}}. \tag{2}$$

We define δ_{ij} as $\frac{e^{-\lambda r(i,j)}}{|\mathcal{B}_i|}$ where $r(i,j)$ denotes the rank of the word b_{ij} in the list $\mathcal{B}_i$. This is essentially an exponential decay function that gives more importance to the top-ranking biasing words for s_i. The hyperparameter α denotes the extent to which $v^*_{s_i}$ is kept close to its corresponding lemma representation v_{s_i}.

Figure 5 shows in a dimensionality-reduced 2D semantic space how the embedding of *digit* is *deconflated* into two of its sense embeddings (anatomical and numerical digit) by placing them in the corresponding regions of the semantic space (occupied by their sense biasing words).

3.1.3 SensEmbed. Sense embeddings can also be induced by applying word embedding techniques to sense-annotated data. SensEmbed [7] is an application of Word2vec in this context. Given the unavailability of large-scale sense-annotated data, we opted for high confidence automatic Word Sense Disambiguation.

3.1.4 Nasari. Nasari [8] is a multilingual vector representation technique which not only enables accurate representation of word senses in different languages, but it also provides two main advantages over existing approaches: (1) high coverage, including both concepts and named entities, (2) comparability across languages and linguistic levels (i.e., words, senses and concepts), thanks to the representation of linguistic items in a single unified semantic space and in a joint embedded space, respectively.

Nasari combines the complementary knowledge derived from the semantic network of BabelNet and statistical information derived from text corpora. The technique analyzes the contents of individual Wikipedia pages, expanded through the semantic relations extracted from BabelNet, using lexical specificity. Lexical specificity is a statistical measure based on the hypergeometric distribution which computes the importance of a given word that occurs in a subset $\mathcal{C}^*$ of a corpus $\mathcal{C}$ as follows:

$$spec(T, t, F, f) = -k \log_e P(X = f) - \log_{10} \left(\sum_{i=f}^{F} a_i \right) \tag{3}$$

T and t are the respective total number of content words in $\mathcal{C}$ and $\mathcal{C}^*$, and F and f are the repective frequencies of the target word in $\mathcal{C}$ and $\mathcal{C}^*$, and k is the natural logarithm of 10 (i.e., $\log_e 10$).

We applied Nasari representations to a series of NLP tasks, including word similarity, sense clustering, domain labeling, and Word Sense Disambiguation, and for each of them reported state-of-the-art performance on several standard datasets across different languages [9]. We also showed that lexical specificty can provide consistent improvements over conventional *tf-idf*.

3.2 Rare and unseen words

The prominent distributional approach to word representation is highly reliant on the availability of large amounts of training data and falls short of effectively modeling rare words that appear only a handful of times in the training corpus. Importantly, the coverage issue is more evident when representations trained on abundant generic texts are applied to tasks in specific domains. As a matter of fact, the target domain can have dedicated lexical resources, such as ontologies, which are generally ignored by the distributional representation approach. We have proposed two different techniques that exploit the knowledge encoded in lexical resources in order to induce representations for rare and unseen words. Our embedding induction approach provides an advantage over the past work in that it enables vocabulary expansion not only for morphological variations, but also for infrequent domain specific terms.

3.2.1 Induction using semantic landmarks. The first technique, proposed in [3] and extended in [10], extracts from a lexical resource a list of *semantic landmarks* $\mathcal{L}_r$ which can best represent a target rare word w_r. Let $\mathbf{d}(x)$ be an embedding for word x in a given word embedding space. A new embedding for w_r in the same semantic space is induced using the following formula:

$$\hat{\mathbf{d}}(w_r) = \theta\, \mathbf{d}(w_r^0) + \frac{1}{\mathcal{L}_r} \sum_{i=1}^{|\mathcal{L}_r|} \mathrm{e}^{-i} \mathbf{d}(l_{i,r}). \tag{4}$$

where $l_{i,r}$ is the i^{th} word in $\mathcal{L}_r$. The formula computes an embedding for w_r which maps the word to the weighted centroid of its semantic landmarks. The exponential weighting assigns more importance to the top words in the list which are semantically more representative of w_r.

3.2.2 Induction using space transformation. Semantic space transformation has been widely studied in the multilingual embedding context. However, the application of these techniques to mono-lingual setting and across heterogeneous semantic spaces has remained largely unexplored. We proposed a methodology that adapts graph em-

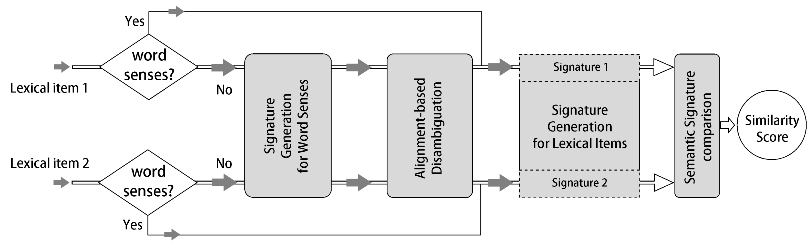

Figure 6
The process of measuring the semantic similarity of a pair of linguistic items using our approach, ADW. A linguistic item is first disambiguated into a set of concepts, if not already sense disambiguated, after which its semantic signature is computed. The similarity of two linguistic items is then calculated by comparing their semantic signatures.

bedding techniques as well as cross-lingual vector space mapping approaches (Least Squares and Canonical Correlation Analysis) in order to merge the distributional and ontological sources of lexical knowledge [4]. Based on this merging we proposed an embedding induction technique for rare and unseen words.

4. Semantic Similarity Measurement

Measuring the extent to which two lexical items (concepts, word senses, words, phrases, etc.) are semantically similar is one of the most popular research fields in lexical semantics, with a wide range of NLP applications.

4.1 Align, Disambiguate, and Walk

By leveraging our semantic representation of word senses (see Section 3.1.1), we put forward an approach [5, 11] that, given a pair of linguistic items, first performs an alignment-based disambiguation and then compares the two at the sense level, independently of their surface form or any ambiguity therein (Figure 6).

4.2 Multilingual

Nasari representations [9] feature multilinguality, with potential application to dozens of languages. This makes the representations readily applicable for similarity measurement in multi-lingual and cross-lingual settings [9]. Also in this path, as a part of the SemEval-2017 Task 2 on Multilingual and Cross-lingual Semantic Word Similarity [12], we provided a reliable framework for the evaluation of multi-lingual and cross-lingual semantic similarity across five languages.

4.3 Cross-level

Semantic similarity has typically been measured across items of approximately similar sizes. As a result, similarity measures have largely ignored the fact that different types of linguistic item can potentially have similar or even identical meanings, and therefore are designed to compare only one type of linguistic item. We introduce in [13, 14] a new semantic evaluation called cross-level semantic similarity (CLSS), which measures the

Table 1
The semantically-aware pseudowords obtained for three polysemous words.

word	type: equivalent pseudoword
arm	forearm*baseball_cap*sword*armchair*executive_branch*garment
plan	retirement_plan*architect*diagram
party	political_party*dinner_party*clique*fiesta*someone

degree to which the meaning of a larger linguistic item, such as a paragraph, is captured by a smaller item, such as a sentence.

4.4 Weighted overlap (WO)

Proposed in [2], Weighted overlap computes the similarity between a pair of ranked lists by comparing the relative rankings of the dimensions. Let H denote the intersection of all non-zero dimensions in positive vectors (or multinomila distributions in general) and $r_h(\mathcal{V})$ be a function returning the rank of the dimension h in the sorted signature $\mathcal{V}$. Weighted overlap calculates the similarity of two vectors $\mathcal{V}_1$ and $\mathcal{V}_2$ as:

$$Sim_{WO}\left(\mathcal{V}_1, \mathcal{V}_2\right) = \frac{\sum_{h \in H} \left(r_h(\mathcal{V}_1) + r_h(\mathcal{V}_2) \right)^{-1}}{\sum_{i=1}^{|H|} (2i)^{-1}} \tag{5}$$

where the denominator is a normalization factor that guarantees a maximum value of one. The measure first sorts the two vectors according to their values and then harmonically weights the overlaps between them. The minimum value is zero and occurs when there is no overlap between the two vectors, i.e., $|H| = 0$. The measure is symmetric and satisfies the top-weightedness property, i.e., it penalizes the differences in the higher rankings more than it does for the lower ones. In a series of experiments we showed that WO can provide consistent improvements over the widely-used cosine distance [5].

5. Word Sense Disambiguation

Word Sense Disambiguation (WSD) is a core research field in computational linguistics, dealing with the automatic assignment of senses to words occurring in a given context. WSD is one of the best examples of NLP tasks which face the so-called knowledge acquisition bottleneck, i.e., the difficulty of capturing knowledge in a computer-usable form. In fact, providing sense-annotated knowledge on a large scale is a time-consuming process, which has to be carried out separately for each word sense, and repeated for each new language of interest.

5.1 Semantically-aware pseudowords

We provided a solution to overcome the knowledge acquisition bottleneck in the evaluation of Word Sense Disambiguation systems [15]. This solution lies on a semantic simulation of polysemous words through *semantically-aware pseudowords*. Each pseudowords

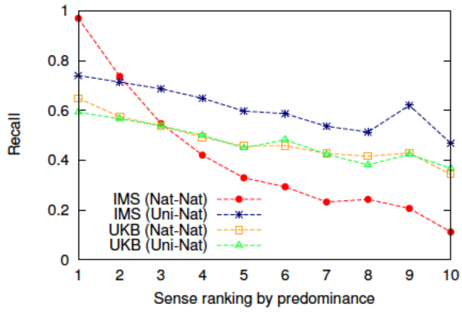

Figure 7
Recall by sense predominance in the naturally distributed test set

is constructed by a set of monosemous words that tend to preserve the semantics of the original word's individual meanings. Table 1 shows three polysemous words and their corresponding semantically-aware pseudowords.

5.2 Large-scale evaluation and analysis of WSD paradigms

We leveraged our semantically-aware pseudowords to create, for the first time, a large-scale evaluation framework for WSD [16]. Using this framework, we are able to perform an experimental comparison of WSD systems on large dataset made up of millions of sense-tagged sentences. Our large-scale framework enables us to carry out an in-depth analysis of the factors and conditions which determine the systems' performance. We provide extensive in-detail experiments on two state-of-the-art WSD systems, IMS (supervised) and UKB (knowledge-based), in various settings (including uniform and natural sense distributions). As an example, Figure 7 shows how system's performance varies by sense predominance, for two different sense distribution settings.

5.3 Embeddings for WSD

In [17], we studied how word embeddings can be used in Word Sense Disambiguation, one of the oldest tasks in Natural Language Processing and Artificial Intelligence. We proposed different methods through which word embeddings can be leveraged in a state-of-the-art supervised WSD system architecture, and performed a deep analysis of how different parameters affect performance. We showed how a WSD system that makes use of word embeddings alone, if designed properly, can provide significant performance improvement over a state-ofthe-art WSD system that incorporates several standard WSD features.

6. Ontology Enrichment and Alignment

Lexical resources and ontologies are repositories of machine-readable knowledge that can be used in virtually any Natural Language Processing task. Notable examples are WordNet, Wikipedia and, more recently, collaboratively-curated resources such as OmegaWiki and Wiktionary. On the one hand, these resources are heterogeneous in design, structure and content, but, on the other hand, they often provide complementary knowledge which we would like to see integrated.

6.1 Alignment of heterogenous lexical resources

Despite being an active field of research, lexical resource alignment techniques have often been specific to a particular pair of resources, or heavily dependent on the availability of hand-crafted alignment data for the pair of resources to be aligned. In [18], we presented a unified approach that can be applied to an arbitrary pair of lexical resources, including machine-readable dictionaries with no network structure. Our approach leverages a similarity measure that enables the structural comparison of senses across lexical resources, achieving state-of-the-art performance on the task of aligning WordNet to three different collaborative resources: Wikipedia, Wiktionary and OmegaWiki.

6.2 Automatic extension of WordNet taxonomy

In [19], we introduced a new resource CROWN that extends the existing WordNet taxonomy, more than doubling the existing number of synsets, and attaches these novel synsets to their appropriate hypernyms in WordNet. CROWN is released in the same format as WordNet and therefore is fully compatible with all existing WordNet-based tools and libraries.

6.3 Ontologizing Wiktionary

In [18], we proposed a technique for transforming arbitrary machine-readable dictionaries to full-fledged ontologies. Our ontologization algorithm takes as input a lexicon L and outputs a semantic graph $G = (V, E)$ where V is the set of concepts in L and E is the set of semantic relations between these concepts. We first create the empty graph $G_L = (V, E)$ such that V is the set of concepts in L and $E = \emptyset$. For each source concept $c \in V$ we create a bag of content words $W = \{w_1, \ldots, w_n\}$ which includes all the content words in its definition d and, if available, additional related words as from lexicon relations (e.g., synonyms in Wiktionary). The problem is then cast as a disambiguation task whose goal is to identify the intended sense of each word $w_i \in W$ according to the sense inventory of L: if w_i is monosemous, i.e., $|\{\mathcal{I}_{G_L}(w_i)\}| = 1$, we connect our source concept c to the only sense c_{w_i} of w_i and set $E := E \cup \{(c, c_{w_i})\}$; else, w_i has multiple senses in L. In that case, we choose the most appropriate concept $c_i \in \mathcal{I}_{G_L}(w_i)$ by finding the maximal similarity between the definition of c and the definitions of each sense of w_i. Having found the intended sense $\hat{c}_{w_i}$ of w_i, we add the edge $(c, \hat{c}_{w_i})$ to E. As a result of this procedure, we obtain a semantic graph representation G for the lexicon L.

7. Statistical Machine Translation

In [20] we reported our first attempt at training an English to Persian phrase-based statistical machine translation system.

7.1 parallel corpora from movie subtitles

Parallel corpora are essential for the training of statistical and neural machine translation systems. We constructed the first large-scale English-Persian parallel corpus by aligning subtitles from hundreds of movies. The corpus, namely TEP, was introduced

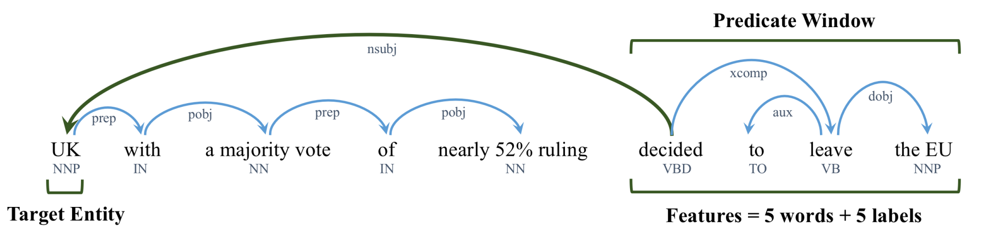

Figure 8
Target context identification for the metonymy candidate “UK”.

in [21]. The resource has played an important role in research towards Persian machine translation.

8. Geographical Parsing

The ability to geo-locate events in textual reports represents a valuable source of information in many real-world applications such as emergency responses, real-time social media geographical event analysis, understanding location instructions in auto-response systems and more. Geoparsing consists of two main sub-problems: identifying and tagging place mentions in a text (called geotagging) and resolving those mentions to their geographical coordinates (called geocoding). Geocoding is a special case of Named Entity Recognition (NER). Geoparsing is still widely regarded as a challenge because of domain language diversity, place name ambiguity, metonymic language and limited leveraging of context as we show in our analysis. In [22] we performed a deep evaluation and analysis of a number of leading geoparsers on different datasets, highlighting the challenges in detail.

8.1 Metonymy detection

Named entities are frequently used in a metonymic manner. They serve as references to related entities such as people and organisations. Accurate identification and interpretation of metonymy can be directly beneficial to various NLP applications, such as Named Entity Recognition and Geographical Parsing. We showed in [23] how a minimalist neural approach combined with a novel predicate window method can achieve competitive results on the SemEval 2007 task on Metonymy Resolution.

8.2 Geocoding

Geocoding is an important step towards building a geographical profile of free-format text documents. Geocoding is a disambiguation problem aiming to resolve a location mention to its coordinates using contextual features to drive the decision. Until now, this task was carried out using disparate methods such as spatial relationships and linguistic cues combined with manually designed heuristics and rules. Our model (called Camcoder) unifies language and space into a single integrated system with no need for further adaptation or expert engineering. It also shows a greater contextual awareness, which enabled it to achieve state-of-the-art performance on multiple datasets as compared to several popular geocoders.

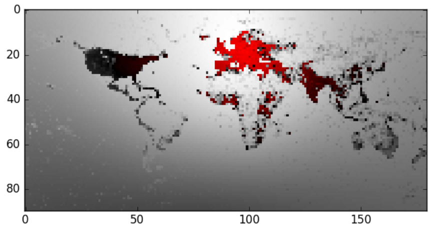

Figure 9
A visual illustration of the spatio-semantic representations of our model and its correlation with the geometric space.

References

1. M. T. Pilehvar, J. Camacho-Collados, R. Navigli, and N. Collier, "Towards a seamless integration of word senses into downstream nlp applications," in *Proceedings of the 55th Conference of the Association for Computational Linguistics*, (Vancouver, Canada), Association for Computational Linguistics, 2017.
2. M. T. Pilehvar, D. Jurgens, and R. Navigli, "Align, Disambiguate and Walk: a Unified Approach for Measuring Semantic Similarity," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, (Sofia, Bulgaria), pp. 1341–1351, 2013.
3. M. T. Pilehvar and N. Collier, "Improved semantic representation for domain-specific entities," in *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, (Berlin, Germany), pp. 12–16, Association for Computational Linguistics, August 2016.
4. V. Prokhorov, M. T. Pilehvar, D. Kartsaklis, P. Lió, and N. Collier, "Learning Rare Word Representations using Semantic Bridging," *ArXiv e-prints*, 2017.
5. M. T. Pilehvar and R. Navigli, "From senses to texts: An all-in-one graph-based approach for measuring semantic similarity," *Artificial Intelligence*, vol. 228, pp. 95–128, 2015.
6. M. T. Pilehvar and N. Collier, "De-conflated semantic representations," in *Proceedings of EMNLP*, (Austin, TX), pp. 1680–1690, 2016.
7. I. Iacobacci, M. T. Pilehvar, and R. Navigli, "Sensembed: Learning sense embeddings for word and relational similarity," in *Proceedings of ACL*, (Beijing, China), pp. 95–105, 2015.
8. J. Camacho-Collados, M. T. Pilehvar, and R. Navigli, "NASARI: a Novel Approach to a Semantically-Aware Representation of Items," in *Proceedings of NAACL*, (Denver, USA), pp. 567–577, 2015.
9. J. Camacho-Collados, M. T. Pilehvar, and R. Navigli, "Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities," *Artificial Intelligence*, vol. 240, pp. 36–64, 2016.
10. M. T. Pilehvar and N. Collier, "Inducing embeddings for rare and unseen words by leveraging lexical resources," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, (Valencia, Spain), pp. 388–393, Association for Computational Linguistics, April 2017.

11. M. T. Pilehvar and R. Navigli, "An open-source framework for multi-level semantic similarity measurement," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, (Denver, Colorado), pp. 76–80, Association for Computational Linguistics, June 2015.
12. J. Camacho-Collados, M. T. Pilehvar, R. Navigli, and N. Collier, "SemEval-2017 task 2: Multilingual and cross-lingual semantic word similarity," in *Proceedings of the 11th International Workshop on Semantic Evaluation*, (Vancouver, Canada), 2017.
13. D. Jurgens, M. T. Pilehvar, and R. Navigli, "SemEval-2014 task 3: Cross-level semantic similarity," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, (Dublin, Ireland), pp. 17–26, 2014.
14. D. Jurgens, M. T. Pilehvar, and R. Navigli, "Cross level semantic similarity: an evaluation framework for universal measures of similarity," *Language Resources and Evaluation*, vol. 50, no. 1, pp. 5–33, 2015.
15. M. T. Pilehvar and R. Navigli, "Paving the way to a large-scale pseudosense-annotated dataset," in *Proceedings of NAACL-HLT*, (Atlanta, USA), 2013.
16. M. T. Pilehvar and R. Navigli, "A large-scale pseudoword-based evaluation framework for state-of-the-art Word Sense Disambiguation," *Computational Linguistics*, vol. 40, no. 4, 2014.
17. I. Iacobacci, M. T. Pilehvar, and R. Navigli, "Embeddings for word sense disambiguation: An evaluation study," in *Proceedings of ACL*, (Berlin, Germany), pp. 897–907, 2016.
18. M. T. Pilehvar and R. Navigli, "A robust approach to aligning heterogeneous lexical resources," in *Proceedings of ACL*, pp. 468–478, 2014.
19. D. Jurgens and M. T. Pilehvar, "Reserating the awesometastic: An automatic extension of the wordnet taxonomy for novel terms," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Denver, Colorado), pp. 1459–1465, Association for Computational Linguistics, May–June 2015.
20. M. Pilevar and H. Faili, "PersianSMT: A first attempt to english-persian statistical machine translation," in *Proceedings of 15th International Conference of Statistical Analysis of Textual Data (JADT)*, (Paris, France), pp. 1101–1112, 2010.
21. M. Pilevar, H. Faili, and A. Pilevar, "TEP: Tehran English-Persian Parallel corpus," in *Computational Linguistics and Intelligent Text Processing*, (Tokyo, Japan), pp. 68–79, 2011.
22. M. Gritta, M. T. Pilehvar, N. Limsopatham, and N. Collier, "What's missing in geographical parsing?," *Language Resources and Evaluation*, pp. 1–21, 2017.
23. M. Gritta, M. T. Pilehvar, N. Limsopatham, and N. Collier, "Vancouver welcomes you! minimalist location metonymy resolution," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, (Vancouver, Canada), 2017.