

## Research and Applications

# Expanding a radiology lexicon using contextual patterns in radiology reports

Bethany Percha,<sup>1,2</sup> Yuhao Zhang,<sup>2</sup> Selen Bozkurt,<sup>3</sup> Daniel Rubin,<sup>4</sup> Russ B Altman,<sup>5,6</sup> and Curtis P Langlotz<sup>4</sup>

<sup>1</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>2</sup>Biomedical Informatics Training Program, Stanford University, Stanford, CA, USA, <sup>3</sup>Department of Biostatistics and Medical Informatics, Akdeniz University Faculty of Medicine, Antalya, Turkey, <sup>4</sup>Department of Radiology, Stanford University, Stanford, CA, USA, <sup>5</sup>Department of Medicine, Stanford University, Stanford, CA, USA and <sup>6</sup>Department of Genetics and Bioengineering, Stanford University, Stanford, CA, USA

Corresponding Author: Bethany Percha, 234 Marshall Street, Suite 10, Redwood City, CA 94063, USA. E-mail: bethany.percha@mssm.edu.

Received 25 July 2017; Revised 1 November 2017; Editorial Decision 12 December 2017; Accepted 18 December 2017

## ABSTRACT

**Objective:** Distributional semantics algorithms, which learn vector space representations of words and phrases from large corpora, identify related terms based on contextual usage patterns. We hypothesize that distributional semantics can speed up lexicon expansion in a clinical domain, radiology, by unearthing synonyms from the corpus.

**Materials and Methods:** We apply word2vec, a distributional semantics software package, to the text of radiology notes to identify synonyms for RadLex, a structured lexicon of radiology terms. We stratify performance by term category, term frequency, number of tokens in the term, vector magnitude, and the context window used in vector building.

**Results:** Ranking candidates based on distributional similarity to a target term results in high curation efficiency: on a ranked list of 775 249 terms, >50% of synonyms occurred within the first 25 terms. Synonyms are easier to find if the target term is a phrase rather than a single word, if it occurs at least 100× in the corpus, and if its vector magnitude is between 4 and 5. Some RadLex categories, such as anatomical substances, are easier to identify synonyms for than others.

**Discussion:** The unstructured text of clinical notes contains a wealth of information about human diseases and treatment patterns. However, searching and retrieving information from clinical notes often suffer due to variations in how similar concepts are described in the text. Biomedical lexicons address this challenge, but are expensive to produce and maintain. Distributional semantics algorithms can assist lexicon curation, saving researchers time and money.

**Key words:** text mining, radiology, lexicons, ontologies, natural language processing

## INTRODUCTION

Information retrieval from clinical documents depends heavily on biomedical lexicons and ontologies, which contain structured information about the entities in a domain, their attributes, and the relationships that connect them.<sup>1,2</sup> Lexicons and ontologies also play a pivotal role

in clinical practice, serving as the underpinnings of structured documentation and enabling interoperability between clinical systems. As more biomedical information becomes available electronically, they will likely form the glue that integrates different software systems and enables large-scale information retrieval and patient phenotyping.<sup>3,4</sup>

However, even the most accurate and complete ontology will fail in clinical settings unless real clinical data, structured and unstructured, can be mapped to concepts and relations from the ontology. For example, a clinical text-mining system must understand that the raw strings “Crohn’s disease,” “Crohn disease,” “regional enteritis,” “Chron disease” (spelling error), and “crohn disease” (capitalization variant) all refer to the same concept. It must also overcome writers’ differences in style and word preference, which can vary by geographic region, subject area, clinical setting, and the individual.

Distributional semantics algorithms,<sup>5,6</sup> which learn vector space representations of words and phrases based on usage patterns in large corpora, can potentially help us identify these variants automatically. Not only would this method dramatically speed up the process of lexicon and ontology expansion, it would also enable us to efficiently create structured lexicons and ontologies for new and unusual clinical domains. Here we apply distributional semantics to the task of expanding RadLex, a manually created lexicon of radiology terms, using distributional information on words and phrases from a corpus of nearly 6 million radiology reports. We discuss the benefits and pitfalls of this approach and develop a set of heuristics to facilitate its application in other clinical domains.

## METHODS

A graphical outline of our methods, including preprocessing, vector creation, and evaluation steps, can be found in Figure 1.

### Radiology corpus: RadCore and STRIDE

Our corpus consisted of the complete RadCore database<sup>7</sup> and a corpus of additional radiology reports from the Stanford Translational Research Integrated Database Environment (STRIDE).<sup>8</sup> RadCore is a multi-institutional database of radiology reports aggregated in 2007 from 3 major health care organizations: Mayo Clinic (812 reports), MD Anderson Cancer Center (5000 reports), and Medical College of Wisconsin (1 893 819 reports). From STRIDE, we added 4056 227 radiology reports of 564 210 patients seen at Stanford Hospital and Clinics since 1998.

Both the RadCore and Stanford radiology report corpora were deidentified by the institutions where they were produced. This retrospective study was approved by the Stanford Institutional Review Board.

### The RadLex lexicon

RadLex is a lexicon of radiology concepts and associated terms created manually over a decade by members of more than 30 professional radiology organizations.<sup>9</sup> The lexicon contains 62 531 terms mapped to 46 037 unique concepts. There are 24 129 unique pairs of terms within RadLex for which both terms map to the same concept (we call these “synonyms”). We parsed the raw RadLex comma-separated file to identify the synonym pairs, as well as all of the parent-child hierarchical relationships in the lexicon.

We recorded the parent category for each unique term in RadLex. Every RadLex term can be traced upward to 1 of 16 general categories: *anatomical entity*, *property*, *procedure*, *procedure step*, *imaging observation*, *clinical finding*, *RadLex descriptor*, *temporal entity*, *obsolete term*, *nonanatomical object*, *nonanatomical substance*, *report component*, *imaging modality*, *process*, *RadLex non-anatomical set*, and *metaclass/other*. In our analysis, we focused

on the 8 most frequent categories represented among the terms in our corpus, which are italicized above.

### Preprocessing: tokenization and concatenation

We used the Stanford CoreNLP toolkit<sup>10</sup> to tokenize the text of every narrative report in our combined corpus (The Stanford tokenizer is free and open source; however, there are also other tokenizers that would be appropriate for this task, such as the *WordPunctTokenizer* from Python’s Natural Language Toolkit.) After lowercasing each token, we wrote the entire corpus to an intermediate text file, with individual tokens separated by spaces. We applied the *word2phrase* tool<sup>11</sup> to this file to concatenate likely phrases. For example, if the words “heterogeneously” and “dense” occurred as the phrase “heterogeneously dense” with high enough frequency, *word2phrase* would concatenate them in the text using an underscore: “heterogeneously\_dense.” We used the default parameters for *word2phrase*.

### Building word and phrase vectors

We used the *word2vec* package<sup>11,12</sup> to build vector representations of all terms in our preprocessed corpus. *Word2vec*’s vectors represent each word or phrase as a mathematical combination of the words and phrases surrounding it within a linear context window (Figure 2). Terms with similar contexts will tend to have similar vectors. The *word2vec* package allows the user to set several parameters, including the vector dimension, the size of the linear context window, and the choice of model (continuous bag-of-words vs skip-gram). We used the skip-gram model with vector dimension 100 (the standard dimension for *word2vec*) and a window width of 1, 3, 5, or 7, and default settings for all other parameters. No vectors were built for terms occurring <5 times in the corpus.

### Synonym retrieval by vector and target term parameters

We performed several experiments to see if *word2vec*’s ability to recognize synonyms varied by vector and/or term properties. For each term in RadLex that (1) had a vector and (2) had a synonym that also had a vector, we ranked all of the corpus terms (with vectors) by cosine similarity to the target term’s vector, and found the position of the synonym on the list. We investigated the position of synonyms with respect to: vector context window width, parent category in RadLex, number of tokens in the target term (whether it was a single word or a phrase), term frequency, and term vector magnitude (measured as its Euclidean, or  $L_2$ , norm). Where applicable, we quantitatively compared distributions of synonym ranks using Kolmogorov-Smirnov tests.

## RESULTS

### RadLex representation in the corpus

*Word2vec* was able to build vectors for 775 248 unique words and phrases in our corpus. Of the 62 531 unique strings in RadLex, 5308 (8.5%) were associated with vectors. A further 187 (0.3%) occurred in the corpus <5 times, so no vectors could be built. The final 57 036 (91.2%) did not occur in the corpus at all. Of the 5308 terms with vectors, 5210 (98.2%) belonged to one of the eight RadLex categories we investigate in this paper. Of the 24 129 unique synonym pairs in RadLex, both terms were associated with vectors in 2383 pairs (9.9%).

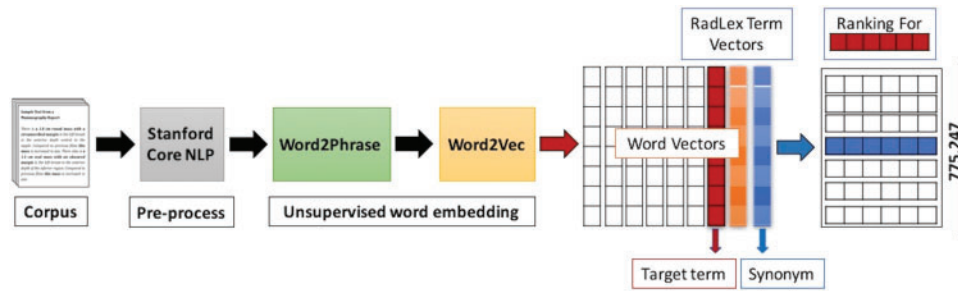


Figure 1. Illustration of our approach to preprocessing, vector creation, and evaluation.

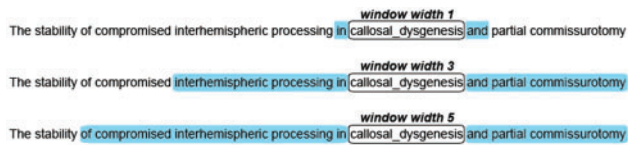


Figure 2. Illustration of the context window sizes for different vector types. The context window for the width-7 vector is not shown.

### Properties of the synonym pairs

For the 906 synonym pairs in which both terms had vectors, the median term frequency for the more frequent term was 3967, while the median term frequency for the less frequent term was 283. There was low correlation between the higher and lower term frequencies (Pearson correlation: 0.30). A very frequent term could have a very infrequent synonym, or a term and its synonym could occur with nearly equal frequency.

### Synonym retrieval by vector window width

Figure 3 shows the percentage of synonyms recovered at each rank cutoff for vectors of varying context window width. On a list 775 247 terms long (1 fewer than the total number of terms with vectors, since the target term was not included), ~50% of synonyms occurred before rank 25. Approximately 75% occurred before rank 2700. Kolmogorov-Smirnov tests revealed no differences between distributions from different window widths ( $P \gg .05$  for all comparisons).

### Synonym retrieval by ontology category

Information about each of the 8 RadLex categories we examined, including example terms, median corpus frequencies, and fractions of terms with vectors, is shown in Table 1. Figure 4 shows the percentage of synonyms recovered by RadLex category. Based on the percentage of synonyms recovered before rank 100, terms from the category *anatomical entity* were distributionally closest to their synonyms; 68.1% occurred before rank 100. Terms from the category *nonanatomical substance* were distributionally farthest from their synonyms, with only 9.1% occurring before rank 100.

### Synonym retrieval by number of tokens

Figure 5 shows the percentage of synonyms recovered by the number of tokens in the target term; that is, whether the term was a word (1 token) or a phrase (2 tokens). Phrases were distributionally more similar to their synonyms than single words; 68.3% of phrases' synonyms were recovered before rank 100, compared to 49.9% of single words' synonyms.

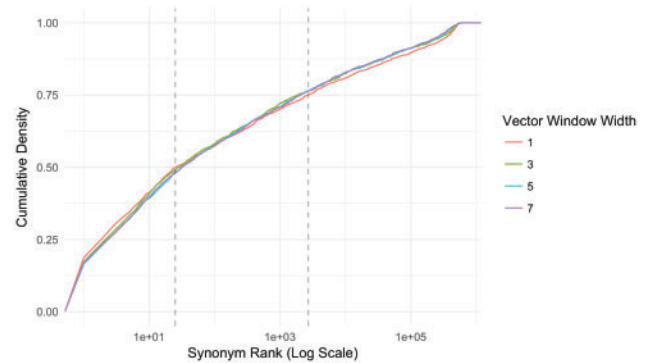


Figure 3. Synonyms recovered by rank cutoff for 4 different vector context window widths.

### Synonym retrieval by term frequency and vector magnitude

Figure 6 shows the percentage of synonyms recovered by the frequency of the target term in the corpus (Figure 6A), and by the magnitude of the target term's vector (Figure 6B). A term will generally be distributionally closer to its synonym the more frequently it occurs in the corpus, although there is little improvement beyond about 1000 occurrences, and, in fact, performance declines slightly for terms with extremely high frequencies. If the target term occurs in the corpus at least 100 times, its synonym will be found within the first 100 terms on the ranked list 63.4% of the time. If the target term occurs <10 times in the corpus, its synonym will only occur within the first 100 terms 8.5% of the time.

As for vector magnitude, synonyms are distributionally closest when the target term's vector magnitude is between 4 and 5. For target terms in this range, 70.8% of synonyms will be found within the top 100 ranked terms. For target terms with very high or low vector magnitude, performance suffers considerably; synonyms will only occur within the top 100 terms 9.1% of the time when the vector magnitude is <2, and 12.5% of the time when the vector magnitude is  $\geq 8$ .

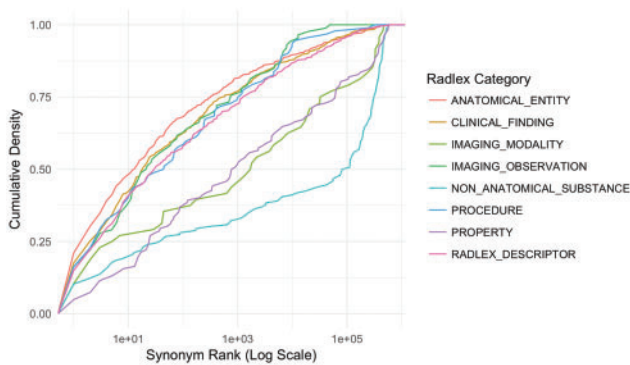
## DISCUSSION

### Interpretation and implications

We observed several key findings that may help guide future lexicon creation efforts. First, the majority of terms in RadLex (91.5%) did not occur in our corpus with sufficient frequency for vector creation. This means that either (1) the majority of RadLex concepts are not discussed in the text of clinical notes, or (2) these concepts *are* referenced in reports, but always in ways that are distinct from those

**Table 1.** Examples of terms from each of 8 RadLex categories

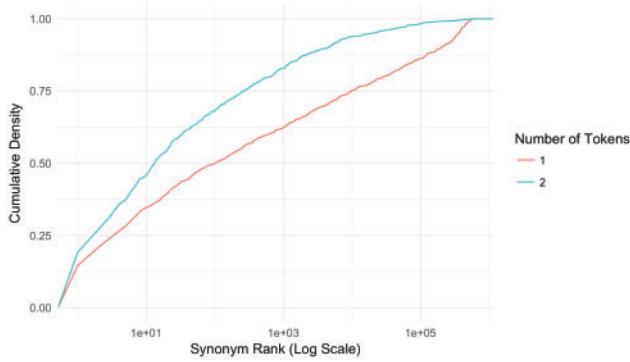
Category	Total terms	Total terms with vectors (%)	Median term frequency in corpus	Example terms
Anatomical entity	51 564	1662 (3.2)	963	<i>bone_marrow</i> <i>pisiform</i> <i>sphenoid_sinus</i> <i>pyloric_sphincter</i> <i>vitreous_chamber</i>
Clinical finding	3106	1360 (43.8)	736	Cholangiohepatitis <i>osteochondroma</i> <i>hepatocellular_carcinoma</i> <i>hemolytic_anemia</i> <i>motor_dysfunction</i>
Imaging modality	125	40 (32.0)	13 173	<i>Qct</i> <i>fluoro</i> <i>spectroscopy</i> <i>mrs</i> <i>tomosynthesis</i>
Imaging observation	1540	181 (11.8)	1125	<i>dromedary_hump</i> <i>linguine_sign</i> <i>corkscrew_collaterals</i> <i>smooth_margin</i> <i>homogeneous_enhancement</i>
Nonanatomical substance	630	280 (44.4)	312	Chlorine <i>zinc</i> <i>lu</i> <i>tc-99m_hsa</i> <i>iodine-131_bippuran</i>
Procedure	856	359 (41.9)	425	Steroid_injection <i>echocardiogram</i> <i>splenectomy</i> <i>renal_autotransplantation</i> <i>jejunoileal_bypass</i>
Property	1807	322 (17.8)	1366	Straining <i>hyperflexion</i> <i>flow</i> <i>inspiration</i> <i>deviation</i>
RadLex descriptor	1747	980 (56.1)	3282	Truncated <i>high_resolution</i> <i>angulated</i> <i>circumscribed</i> <i>sharply-defined</i>

**Figure 4.** Synonyms recovered by rank cutoff for 8 different RadLex categories.

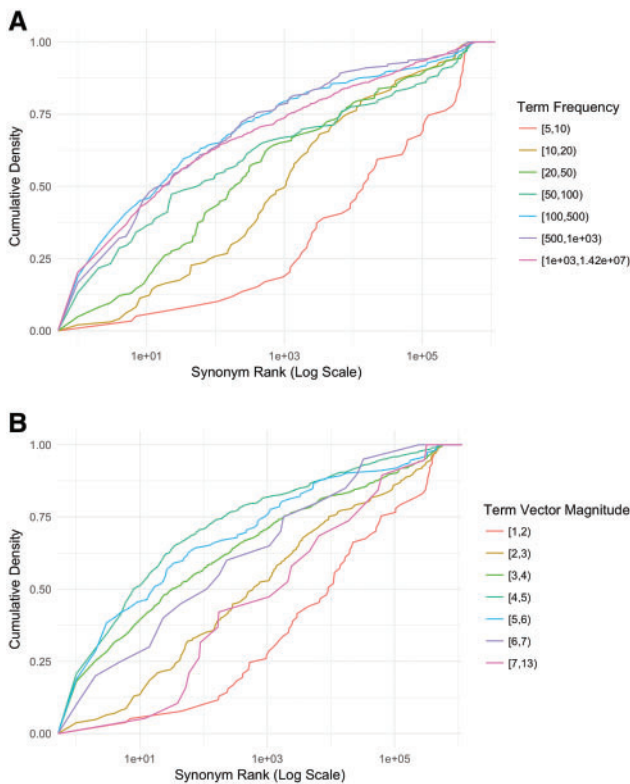
listed in RadLex. Since RadLex was built specifically for the task of streamlining radiology reporting in the context of clinical documentation, we find the second interpretation more likely.

Unfortunately, the second interpretation also highlights the major weakness of distributional approaches: one must always start with a known target term of interest that also has a vector. If the target term is unknown (ie, if you want to add terms for a brand-new concept to the lexicon) or does not have a vector, these approaches are of limited utility. For example, the terms *medial\_intercondylar\_eminence\_of\_tibia* and *interlobar\_vein\_of\_right\_kidney* are both RadLex terms without vectors. It is possible that (1) these concepts are never described in the reports from our corpus, or (2) they are simply described so inconsistently that no pattern occurs frequently enough for a vector to be built for them. One could address (1) by simply gathering a larger corpus, but (2) is challenging even when the corpus is large.

Second, we found only weak correlation between the frequencies of terms and their synonyms. However, on average, the less frequent synonyms tended to be about 7% as frequent as the more frequent synonyms in the corpus. If we assume that the more frequent terms are the search terms and that terms and their synonyms are used in distinct sets of documents (eg, from 2 different institutions), we can



**Figure 5.** Synonyms recovered by rank cutoff for terms 1 token long (words) and 2 tokens long (phrases).



**Figure 6.** (A) Synonyms recovered by rank cutoff for terms with different frequency of occurrence in the corpus. (B) Synonyms recovered by rank cutoff for terms with vectors of different magnitude.

estimate that, on average, we would retrieve roughly 7% more documents for each synonym we identify. This could make a material difference for many applications, and highlights the need for efficient methods of identifying likely synonyms.

Upon beginning the search for synonyms, one must start with a target term. There are several features of the target term that indicate a greater chance of success in the synonym search. High term frequency (>100 occurrences in the corpus) and a target term that is a phrase rather than a single word will lead to greater success identifying synonyms using distributional approaches. Vector magnitude, a measure both of term frequency and of the consistency of a term's context,<sup>13</sup> is also an indicator of likely success: intermediate vector magnitudes are optimal, indicating terms that occur quite

frequently, but not so frequently that they are dispersed over a wide variety of different contexts (eg, “the,” “it”).

The term's category also plays a role. In the case of RadLex, it was easier to recover synonyms for target terms that were anatomical entities, clinical findings, and imaging observations than it was to recover synonyms for properties and nonanatomical substances, even though the median term frequencies among the various categories did not differ substantially (Table 1). Anatomical terms and disease names tend to be very specific and used in specific contexts, whereas properties like “flow” and “inspiration” and nonanatomical substances like “chlorine” and “zinc” have a variety of biological meanings and can be used in several different contexts. We suspect this is the source of the discrepancy.

Finally, and surprisingly, we observed virtually no difference in performance when we used vectors built with different context window sizes. It appears that most of the distributional information that allows us to identify synonyms for radiology terms occurs within the words immediately preceding and following the target term.

### Some challenges for distributional lexicon learning

Some examples of ranked term lists for three different RadLex target terms are shown in Table 2. It is immediately obvious that very few pairs of distinct biomedical terms are actually genuine synonyms. Instead, what distributional approaches produce are terms that are used in similar contexts, which can include highly related terms that are not true synonyms (eg, two different joints) or even antonyms. This highlights the need for manual review of all findings. However, it could also be seen as a positive feature for broader lexicon curation. Since the curation process we describe is likely to yield new term candidates that are used contextually in ways similar to existing lexicon terms, it may be possible to discover brand-new lexicon concepts with this approach.

As we can see from the term “heterogeneous” in Table 2, the concatenation of multiword terms also represents a potential issue for distributional methods, since separate vectors are often built for a phrase and the individual words within that phrase. What's more, each occurrence of the phrase “heterogeneously dense,” for example, will contribute to only one vector: the “heterogeneously\_dense” vector if that occurrence happens to be concatenated, and the two individual word vectors if it is not. All three vectors are distinct, and the vector for “heterogeneously\_dense” is not a simple mathematical combination of the other two. This is a problem for distributional approaches in general and an active area of research.

Finally, lexicon expansion (the task we address here) is a different, and in some ways simpler, task than actually *identifying* lexicon terms in a new corpus. The latter task would require, in addition to the lexicon itself, a set of rules that address issues like word sense disambiguation; the noun “test” and the verb “test” share a vector, for example, but perhaps we only want the noun. There are ways this might be approached from a distributional perspective, but they are beyond the scope of this paper.

### Related work in biomedical lexicon and ontology learning

Our work builds on decades of former work in biomedical text mining, mostly within the field of biomedical named entity recognition and normalization.<sup>14–18</sup> Several authors have investigated which features provide the best performance in biomedical named entity recognition, including distributional features.<sup>19,20</sup> We also draw

**Table 2.** The 10 closest corpus terms to 3 example RadLex terms

RadLex Term	Synonym candidate	Comment
hyperemia	hyperemia	* misspelling
	hyperemic	*
	wall-thickening	symptom
	pericholecystic_fluid	*
	blood_flow	* misspelling
	increased_flow/hyperemia	*
	hyperemia	symptom
	wall_hyperemia	
	pericholecystic_fluid	
	wall_thickening	
carpometacarpal_joint	interphalangeal_joint	different joint
	carpal_metacarpal	*
	metacarpophalangeal_joint	different joint
	dip_joint	"
	first_carpometacarpal	*
	carpal-metacarpal_joint	*
	carpometacarpal_joints	*
	third_dip	different joint
	ip_joint	"
	second_mcp	"
heterogeneous	heterogeneous	* spelling variant
	somewhat_heterogeneous	*
	heterogeneously	*
	inhomogeneous	*
	mildly_heterogeneous	*
	markedly_heterogeneous	*
	heterogeneously	*
	somewhat_heterogeneous	*
	lobulated	error
	diffusely_heterogeneous	*

Asterisks indicate terms that are synonymous with the RadLex terms or more specific versions of them.

heavily on previous work in biomedical ontology learning.<sup>21-23</sup> The problem of recognizing biomedical synonyms and normalizing them to database identifiers automatically was attacked head-on by the biomedical natural language processing community in the BioCreative competitions.<sup>24,25</sup> Our work expands on these approaches by applying them to a new domain (radiology) and by considering the problem of lexicon expansion in a practical, curator-oriented context.

### Related work in clinical text annotation

Ultimately, the goal of building a lexicon for a domain such as radiology is to extract structured information from the unstructured text of clinical documents. There are alternative approaches to this task that do not start from lexicons, although many clinical information extraction systems incorporate lexicons within larger rule-based or statistical frameworks. Examples of such systems include MedLee,<sup>26</sup> cTakes,<sup>27</sup> and MetaMap<sup>28</sup>; these have been compared head to head on at least one task.<sup>29</sup>

### CONCLUSIONS AND FUTURE WORK

Distributional approaches represent a practical and principled way to approach lexicon curation and expansion. Although they still require curators to manually review term lists, this is preferable to more ad hoc approaches, and it captures many unusual spelling and concatenation variants that a human might not think of on his or

her own. We have developed several practical heuristics for lexicon building using this approach in the radiology domain, most of which we have kept largely qualitative in an effort to assist lexicon creation across multiple domains.

We hope that our work inspires others to apply distributional methods to assist a variety of curation tasks.

### FUNDING STATEMENT

This work was supported by National Institutes of Health grant numbers U01CA142555, 1U01CA190214, 1U01CA187947, LM05652, and GM61374, and National Institute of Biomedical Imaging and Bioengineering subcontract number HHSN2682015000247A. BP was supported by a Morgridge Family Stanford Interdisciplinary Graduate Fellowship.

### COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

### CONTRIBUTORS

BP, YZ, and SB performed the preprocessing and experiments, created the figures, and drafted the manuscript. BP, CL, RBA, and DR jointly conceived of the idea for the paper. CL and DR provided the raw datasets and RadLex lexicon, and interpreted the results from

the perspective of practicing radiologists. RBA edited the manuscript and advised BP and YZ throughout. All authors were involved in the preparation of the final manuscript.

## REFERENCES

- Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinform*. 2008;9(1):75–90.
- Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform*. 2008;67–79.
- Oellrich A, Collier N, Groza T, et al. The digital revolution in phenotyping. *Brief Bioinform*. 2016;17(5):819–30.
- Harispe S, Ranwez S, Janaqi S, Montmain J. The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. *Bioinformatics*. 2014;30(5):740–42.
- Turney PD, Pantel P. From frequency to meaning: vector space models of semantics. *J Artif Intell Res*. 2010;37:141–88.
- Cohen T, Widdows D. Empirical distributional semantics: methods and biomedical applications. *J Biomed Inform*. 2009;42(2):390–405.
- Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. *Artif Intell Med*. 2016;66:29–39.
- Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE: An integrated standards-based translational research informatics platform. In *AMIA Annu Symp Proc*. 2009:391–95.
- Langlotz, CP. RadLex: a new method for indexing online educational materials. *RadioGraphics*. 2006;26(6):1595–97.
- Manning CD, Surdeanu M, Bauer J, et al. The Stanford CoreNLP natural language processing toolkit. In *ACL (System Demonstrations)*. 2014: 55–60.
- Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 2013: 3111–3119.
- Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. 2013.
- Schakel AMJ, Wilson BJ. Measuring word significance using distributed representations of words. *arXiv preprint arXiv:1508.02297*. 2015.
- Jimeno A, Jimenez-Ruiz E, Lee V, et al. Assessment of disease named entity recognition on a corpus of annotated sentences. *BMC Bioinformatics*. 2008;9(3):S3.
- Krallinger M, Leitner F, Rabal O, et al. CHEMDNER: The drugs and chemical names extraction challenge. *J Cheminform*. 2015;7(Suppl 1):S1.
- Rocktäschel T, Weidlich M, Leser U. ChemSpot: a hybrid system for chemical named entity recognition. *Bioinformatics*. 2012;28(12): 1633–40.
- Leaman R, Gonzalez G. BANNER: an executable survey of advances in biomedical named entity recognition. *Pac Symp Biocomput*. 2008;13:652–63.
- Settles B. ABNER: an open source tool for automatically tagging genes, proteins and other entity names in text. *Bioinformatics*. 2005;21(14): 3191–92.
- Munkhdalai T, Li M, Batsuren K, et al. Incorporating domain knowledge in chemical and biomedical named entity recognition with word representations. *J Cheminform*. 2015;7(Suppl 1):S9.
- Tang B, Cao H, Wang X, et al. Evaluating word representation features in biomedical named entity recognition tasks. *BioMed Res Int*. 2014:Art. ID 240403.
- Liu K, Hogan WR, Crowley RS. Natural language processing methods and systems for biomedical ontology learning. *J Biomed Inform*. 2011;44(1):163–79.
- Ruiz-Martínez JM, Valencia-García R, Fernández-Breis JT, et al. Ontology learning from biomedical natural language documents using UMLS. *Expert Syst Appl*. 2011;38(10):12365–78.
- Valencia-García R, Fernández-Breis JT, Ruiz-Martínez JM, et al. A knowledge acquisition methodology to ontology construction for information retrieval from medical documents. *Expert Syst*. 2008;25(3): 314–34.
- Krallinger M, Vazquez M, Leitner F, et al. The Protein-Protein Interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. *BMC Bioinform*. 2011;12(Suppl 8):S3.
- Morgan AA, Lu Z, Wang X, et al. Overview of BioCreative II gene normalization. *Genome Biol*. 2008;9(Suppl 2):S3.
- Friedman C, Alderson PO, Austin JH, et al. A general natural language text processor for clinical radiology. *J Am Med Inform Assoc*. 1994;1(2): 161–74.
- Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical text analysis and knowledge evaluation system (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010;17(5):507–13.
- Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. 2010;17(3): 229–36.
- Wu Y, Denny JC, Rosenbloom ST, et al. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries. *AMIA Annu Symp Proc*. 2012;997–1003.