Automatic Extraction of Linguistic Metaphor with LDA Topic Modeling

Ilana Heintz*, Ryan Gabbard*, Mahesh Srinivasan+, *, David Barner+, Donald S. Black*, Marjorie Freedman*, Ralph Weischedel*

* Raytheon BBN Technologies
10 Moulton St,
Cambridge MA 02139

{iheintz, rgabbard, mfreedman, dblack, rweischedel}@bbn.com

+University of California, San Diego
5336 McGill Hall,
9500 Gilman Drive
La Jolla, CA 92093-0109

barner@ucsd.edu, mahesh.srinivasan@gmail.com

Abstract

We aim to investigate cross-cultural patterns of thought through cross-linguistic investigation of the use of metaphor. As a first step, we produce a system for locating instances of metaphor in English and Spanish text. In contrast to previous work which relies on resources like syntactic parsing and WordNet, our system is based on LDA topic modeling, enabling its application even to low-resource languages, and requires no labeled data. We achieve an F-score of 59% for English.

1 Introduction

Patterns in the use of metaphors can provide a great deal of insight into a culture. Cultural differences expressed linguistically as metaphor can play a role in matters as complex and important as diplomatic relations. For instance, Thornborrow (1993) discusses the different metaphors that are used in the context of *security* in French and British coverage of two major post-cold-war summit meetings. Example metaphors such as "the cornerstone of the new security structure," "structures for defence and security cooperation," and "the emerging shape of Europe," exemplify the English use of the **source concept** *structure* in describing the **target concept** of *security*. In contrast, the metaphors "des règles de sécurité nouvelles (new rules of security)", "une révision fondamentale des dispositions de sécurité (a fundamental revision of security provisions)", and "un système de sécurité européen (a system of European security)" exemplify the French use of the more abstract source concept *system* to describe the same target concept. As Thornborrow notes, the implied British conception of security as "concrete, fixed, and immobile" contrasts deeply with the French conception of security as "a system as a series of processes."

Our ultimate goal is to use metaphor to further our knowledge of how different cultures understand complex topics. Our immediate goal in this paper is to create an automated system to find instances of metaphor in English and Spanish text.

Most existing work on metaphor identification (Fass, 1991; Martin, 1994; Peters and Peters, 2000; Mason, 2004; Birke and Sarkar, 2006; Gegigan et al., 2006; Krishnakumaran and Zhu, 2007; Shutova et al., 2010; Shutova et al., 2012)[1] has relied on some or all of handwritten rules, syntactic parsing, and semantic databases like WordNet (Fellbaum, 1998) and FrameNet (Baker et al., 1998). This limits the approaches to languages with rich linguistic resources. As our ultimate goal is broad, cross-linguistic application of our system, we cannot rely on resources which would be unavailable in resource-poor languages. Instead, we apply LDA topic modeling (Blei et al., 2003b) which requires only an adequate amount of raw text in the target language. This work is similar to Bethard et al. (2009), in which an SVM model is trained with LDA-based features to recognize metaphorical text. There the work is framed as a classification task, and supervised methods are used to label metaphorical and literal text. Here, the task is one of recognition, and we use heuristic-based, unsu-

[1] See Shutova (2010) for a survey of existing approaches

Proceedings of the First Workshop on Metaphor in NLP, pages 58–66,
Atlanta, Georgia, 13 June 2013.

pervised methods to identify the presence of metaphor in unlabeled text. We hope to eliminate the need for labeled data which, as discussed in Bethard et al. (2009) and elsewhere, is very difficult to produce for metaphor recognition.

2 Terminology

We will refer to a particular instance of metaphorical language in text as a **linguistic metaphor**. Each such metaphor talks about a **target concept** in terms of a **source concept**. For example, in "Dems, like rats, will attack when cornered" the source concept is *animals* and the target concept is *politicians*[2], or at a higher level, *governance*. The abstract mapping between a source concept and a target concept will be referred to as a **conceptual metaphor** which is **grounded** by a collection of linguistic metaphors.

In this work, we restrict our attention to a single target concept, *governance*. Our definition of *governance* is broad, including views of the governed and those who govern, institutions of government, laws, and political discourse. We used a large collection (see **Table 1**) of potential source concepts. Beginning with the source concepts of **primary metaphors**, which are hypothesized to be universal (Grady, 1998), we expanded our set to include source concepts commonly found in the scientific literature about metaphor, as well as those found by human annotators manually collecting instances of governance-related metaphors.

Animals	Fishing	Plants
Baseball	Flight	Race
Body	Football	Religion
Botany	Gambling	Sick
Boundary	Grasp	Size
Chess	Health	Sound
Color	Height	Sports
Combustion	Light	Taste
Cooking	Liquid	Temperature
Courtship	Machine	Texture
Cut	Maritime	Theater
Directional force	Money	Time of day
Dogs	Motion	Toxicity
Drug use	Mythology	Vehicle
Electricity	Natural disasters	War
Energy source	Nuclear	Weaponry
Entry	Odor	Weather
Family	Pathways	Weight
Farming	Physical structure	Wild west
Fight	Planning	

Table 1: English Source Concepts

3 High-level system overview

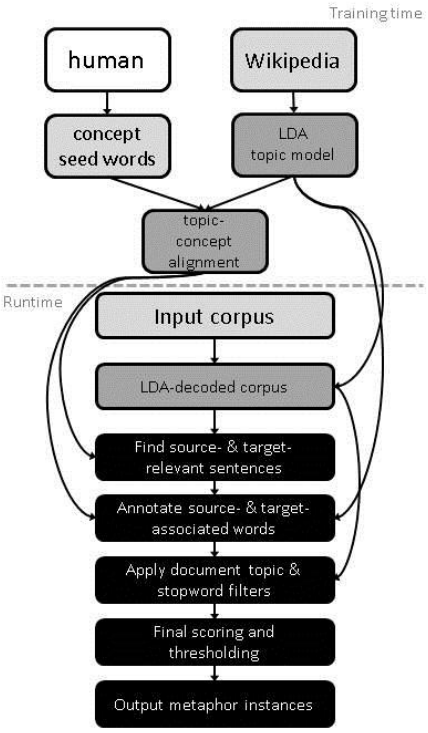

Figure 1: System Overview

Our main hypothesis is that metaphors are likely to be found in sentences that exhibit evidence of both a source and a target concept. The core idea of our system is to use LDA topics as proxies for semantic concepts which may serve as the source or target for a metaphor. For a given language, we build an LDA model from Wikipedia and then align its topics to potential source and target concepts, which are defined by small human-created lists of seed words.

At runtime, the system first does LDA inference on our input corpus to get topic probabilities for each document and sentence. The system then selects those sentences linked by LDA to both a source-aligned topic and a target-aligned topic.[3] For example, a sentence containing "*…virtud so-*

[2] "Dems"" refers to the Democratic Party, an American political party

[3] This is a distant, automatic relative of the 'directed-search' technique of Martin (1994).

cial para construir la democracia…"[4] will be selected because LDA strongly associates it with both the topic [*elecciones, ministro, sucesor, …*][5], aligned to the target concept *governance*, and the topic [*edificio, arquitectura, torre,…*] [6], aligned to the source concept *physical structure.*

Next, the system identifies the words in each selected sentence that are strongly associated with each concept. In the sentence above, it marks *virtud* and *democracia* as target-associated and *construir* as source-associated.

Next it applies two filters. First, we exclude any sentence with too few words that are not LDA stopwords, because the model's predictions may be very inaccurate in these cases. Second, if the topic associated with the source model for a sentence is also a top-ranked topic for the document as a whole, the sentence is excluded. The reason for this is that if the source concept is present throughout the document, it is probably being used literally (see **Figure 2**).

Finally, it uses previously-computed information to determine a final score. All linguistic metaphors scoring above a certain threshold are returned. By varying this threshold, the user can vary the precision-recall tradeoff as needed. A diagram of the system can be found in **Figure 1**.

> Our county has many roads in bad shape. Thousands of our bridges are structurally deficient. Congress needs to pass a new highway bill.

Figure 2: Even though the last sentence is relevant to the source concept *pathways* and the target concept *governance*, it will be correctly rejected because *pathways*-aligned topics are present throughout the document.

4 Implementation Details: Training

Our runtime system requires as input an LDA model, a list of seed words for each concept, and an alignment between concepts and LDA topics.

4.1 LDA Topic Model

The topics defined by LDA topic modeling serve as stand-ins for the more abstractly-defined source and target concepts underlying the metaphors. The input to training our LDA model is the full text of Wikipedia articles in the target language. Wikipedia is available in numerous languages and serves as a corpus of general knowledge, providing us with topics corresponding to a broad range of concepts. Our LDA model is trained using MALLET (McCallum, 2002) for 1000 iterations with 100 topics, optimizing hyperparameters every 10 iterations after a 100 iteration burn-in period. The 500 most common tokens in the training corpus were used as stopwords. The result of LDA is 100 topics, where each topic is a probability distribution over the training corpus vocabulary. Representative words for example English topics are shown in **Figure 3**.

> theater stage musical miss actreess
> theory philosophy pp study scientific knowledge
> nfl bowl yards coach players card yard
> governor republican senate election congress

Figure 3: Sample LDA topics with representative terms

4.2 Concept Seed Word Lists

For each concept c, we have a label and a small set of *seed words* representing that concept, referred to as $K(c)$. These lists were created by hand in English and then translated into Spanish by native speakers. The translation was not intended to be exact; we instructed the annotators to create the lists in a way that was appropriate for their language and culture. For instance, the *football* topic for English describes American football, but in Spanish, the same topic describes soccer.

4.3 Concept-Topic Alignment

The final input to our system is an alignment between concepts and topics, with every topic being mapped to at most one concept. In addition to the seed lists and LDA model, this alignment process takes a score threshold z_{align} and a maximum number of alignments per source and target concept N_S and N_T.

The alignment algorithm is as follows. We align each topic t to the concept c with the maximum score $\lambda(c,t)$, which measures the concept terms' summed probability in the LDA topic: $\lambda(c,t) = \sum_{w \in K(c)} p(w|t)$. We remove all alignments where $\lambda(c,t) < z_{align}$. Finally, for each concept, only the N highest scoring alignments are kept, where N may be different for source and

[4] social virtue to build democracy
[5] elections, minister, successor
[6] building, architecture, tower

target. We refer to the aligned topics for a concept c as $\Lambda(c)$.

Label	Seed List Words	Aligned Topics
Vehicle	***vehicle, wheels, gas, bus***	***0.035: engine, car, model***
		0.29: railway, trains, train
		0.022: energy, gas, linear
Animals	animal, beast, cattle	0.066: animals, animal, species
Courtship	courtship, romance, court	None
Governance	aristocrat, bipartisan, citizen, duke	0.25: Election, elected, parliament
		0.22: Governor, republican, Senate
		0.14: sir, lord, henry
		0.13: kingdom, emperor, empire
		0.12: rights, legal, laws

Table 2: Sample concepts, manually-created seed lists, and aligned topics

A last condition on the topic-concept alignment is the assignment of topics to *trump concepts.* Our only trump concept in this study is *war*. If an LDA topic is aligned with both the *war* concept and the *governance* concept, it is removed from alignment with the *governance* concept. We do this because *war* is so tightly associated with governments that the alignment algorithm invariably aligns it to the *governance* topic. However, *war* is also a very important source concept for *governance* metaphors; our choice is to suffer on recall by missing some governance-relevant sentences, but increase recall on metaphors for which the source concept is *war*. Sample topic-concept alignments are shown in**Table 2**. By inspecting the resulting alignments by hand, we chose the following parameter values for both languages: z_{align}=0.01, N_S=3, N_T=5.

The process of defining concepts is simple and fast and the alignment method is inexpensive. Therefore, while we have not captured all possible source concepts in our initial list, expanding this list is not difficult. We can define new source concepts iteratively as we analyze metaphors that our extraction system misses, and we can add target concepts as our interests broaden.

5 Implementation Details: Runtime

The system receives as input a corpus of documents, their LDA decodings, the LDA decodings of each sentence treated as a separate document, and the topic-concept alignments. Each four-tuple (L, S, T, x) is processed independently, where L is the language, S is the source concept, T is the target concept, and x is the sentence.

Determining Concept Relevance: Recall our basic intuition that a sentence relevant both to an LDA topic in $\Lambda(S)$ (termed **source-relevant**) and one in $\Lambda(T)$ (termed **target-relevant**) is potentially metaphorical. The system judges a sentence x to be C-relevant if the probability of C-aligned topics in that sentence is above a threshold: $\rho_C(x) = \sum_{t \in \Lambda(C)} p(t|x) \geq z_{rel,C}$, where $z_{rel,C}$ is an adjustable parameter tuned by hand. $z_{rel,S}$ is 0.06 in English and 0.05 in Spanish. $z_{rel,T}$ is 0.1 in both languages. On the source side, the system removes all topics in $\Lambda(T)$ from $p(t|x)$ and renormalizes before determining relevance in order to avoid penalizing sentences for having very strong evidence of relevance to governance in addition to providing evidence of relevance to a source concept. For reference below, let $\rho_{ST}(x) = \rho_S(x)\rho_T(x)$ (a measure of how strongly the sentence is associated with its topics) and let $R_C(x) = argmax_{t \in \Lambda(C)}\, p(t|x)$ (the most probable C-aligned topic in the sentence).

If x is not both source- and target-relevant, the system stops and the sentence is not selected.

Finding Concept-Associated Words: The system next creates sets A_C of the words in x associated with the concept C. Let $\sigma_C(w) = \sum_{t \in C}\, p(t|x)$. Then let $A'_C = \{w \in x \,|\sigma_c(w) \geq z_{word}\}$, where z_{word} is a hand tuned parameter set to 0.1 for both languages. That is, any word whose probability in the topic is higher than a theshold is included as a concept-associated word in that sentence. Let $A_S = A'_S - A'_T$ and vice-versa. Note that words which could potentially be associated with either concept are associated with neither. For reference below, let $\omega_c(w) = max_{w \in x}\, \sigma_C(w)$ (the most strongly concept-associated words in the sentence)

and $\omega_{ST}(x) = \omega_S(x)\omega_T(x)$ (the combined strength of those associations).

If x lacks words strongly associated with the source and target concepts (that is, A_S or A_T is empty), the system stops and the sentence is not selected.

Filters: The system applies two filters. First, x must have at least four words which are not LDA stopwords; otherwise, the LDA predictions which drive the system's concept-relevance judgements tend to be unreliable. Second, the most likely source topic $R_S(x)$ must not be one of the top 10 topics for the document as a whole, for reasons described above. If either of these requirements fail, the system stops and the sentence is not selected.

Final Scoring: Finally, the system determines if

$$\ln\left(\lambda_S(R_S(x))\lambda_T(R_T(x))\rho_{ST}(x)\omega_{ST}(x)\right) > z_{final}$$

where z_{final} is a hand-tuned threshold set to -10.0 for English and -13.0 for Spanish. This takes into account the strength of association between topics and the sentence, between the annotated words and the topics, and between the topics and their aligned concepts. Any sentence passing this threshold is selected as a linguistic metaphor.

6 Example Output

We provide examples of both true and false positives extracted by our system. The annotations of source and target-associated words in each sentence are those defined as A_S and A_T above. The source concept *animals* is used for all examples.

1. **Moderates**$_T$ we all hear are an **endangered**$_S$ **species**$_S$, Sen. Richard
2. **Dems**$_T$ like **rats**$_S$ sometimes attack when cornered
3. **Obama**$_T$'s world historical political ambitions **crossbred**$_S$ with his
4. At least **Democratic**$_T$ **representatives**$_T$ are **snakehead**$_S$ fish
5. Another **whopper**$_S$ from Cleveland, **GOP**$_T$ lawyer backs him up
6. Previous post: Illinois **GOP**$_T$ **lawmaker**$_T$ arrested in **animal**$_S$ feed bag related incident
7. Next post: National Enquirer catfighting **Michelle Obama**$_T$ has **claws**$_S$ out for that nice Ann Romney
8. Sen. Lisa **Murkowski**$_T$ R AK independent from Alaska - thank you silly Repubs, **tea**$_S$ party her out ha

Examples 1 through 4 are correct metaphors extracted by our system. In each, some words related to the target concept *governance* are described using terms related to the source concept *animals*. Example 1 best represents the desired output of our system, such that it contains a governance- and animals-relevant metaphor and the terms associated with the metaphor are properly annotated. Some issues do arise in these true positive examples. Example 2, while often termed a simile, is counted as a metaphor for our purposes. In example 3, the source term is correctly annotated, but the target terms should be *political ambitions* rather than *Obama*. It is unclear why the term *snakehead* but not the term *fish* in example 4 is associated with the source concept.

Examples 5 through 8 represent system errors. In example 5, the fact that the word *whopper* occurs frequently to describe a large animal (especially a fish) causes the sentence to be mistakenly identified as relevant to the source concept *animal*. The source term *animal* in example 6 is clearly relevant to the source concept, but it is being used literally. The document-level source concept filtering does not entirely eliminate this error class. While example 7 contains a metaphor and has some relationship to American politics, it would be counted as an error in our evaluations because the metaphor itself is not related to *governance*. In example 8, we have two errors. First, *tea* is strongly present in the topic aligned to the *animal* concept, causing the sentence to be incorrectly marked as source-relevant. Second, because our topic model operates at the level of individual words, it was unable to recognize that *tea* here is part of the fixed, *governance*-related phrase *tea party*.[7]

7 Evaluation

7.1 Collecting Evaluation Data

We collected a domain-specific corpus in each language. We curated a set of news websites and governance-relevant blogs in English and Spanish and then collected data from these websites over the course of several months. For each language, we ran our system over this corpus (all steps in

[7] an American political movement

Section 5), produced a set of linguistic metaphors for each topic-aligned source concept (the target concept was always *governance*), and ranked them by the final score (Section 4.4). Below, we will refer to the set of all linguistic metaphors sharing the same source and target concept as a conceptual metaphor.

7.2 Simple Evaluation

For this evaluation, we selected the top five examples for each conceptual metaphor. If the same sentence was selected by multiple conceptual metaphors, it was kept for only the highest scoring one. We then added enough of the highest-ranked unselected metaphors to create a full set of 300. We then added random sentences from the corpus that were *not* selected as metaphorical by the system to bring the total to 600. Our Spanish annotators were unavailable at the time this evaluation took place, so we are only able to report results for English in this case.

For each of these instances, two annotators were asked the question, "Is there a metaphor about governance in this example?" These annotators had previous experience in identifying metaphors for this study, both by searching manually in online texts and evaluating previous versions of our system. Over time we have given them feedback on what does and does not constitute a metaphor. In this case, the annotators were given neither the system's concept-word association annotations nor the source concept associated with the instance. In one way, the evaluation was generous, because any metaphor in the extracted sentence would benefit precision even if it was not the metaphor found by our system. On the other hand, the same is true for the random sentences; while the system will only extract metaphors with source concepts in our list, the annotators had no such restriction. This causes the recall score to suffer. The annotation task was difficult, with a κ-score of 0.48. The resulting scores are given in **Table 3**. The examples given in Section 5 illustrate the error classes found among the false positives identified by the human annotators. There are many cases where the source-concept associated terms are used literally rather than metaphorically, and many cases where the system-found metaphor is not about governance. Some text processing issues, such as a bug in our sentence breaking script, as well as the noisy nature of blog and blog comment input, caused some of the examples to be difficult to interpret or evaluate.

Annotator	Precision	'Recall'	F	Kappa
1	65	67	66	0.48
2	43	60	50	
Mean	54	64	59	

Table 3: Simple English Evaluation

7.3 Stricter Evaluation

Common Experimental Setup

We did a second evaluation of both English and Spanish using a different paradigm. For each language, we selected the 250 highest-ranked linguistic metaphor instances in the corpus. Subjects on Amazon Mechanical Turk were shown instances with the system-predicted concept-associated words highlighted and asked if the highlighted words were being used metaphorically (options were *yes* and *no*). Each subject was randomly asked about roughly a quarter of the data.

We paid the subjects $10 per hour. We added catch trial sentences which asked the subject to simply answer *yes* or *no* as a way of excluding those not actually reading the sentences. Subjects answering these questions incorrectly were excluded (17 in English, 25 in Spanish).

We defined the **metaphoricity** of an instance to be the fraction of subjects who answered *yes* for that instance. We define the metaphoricity of a conceptual metaphor as the average metaphoricity of its groundings among the instances in this evaluation set.

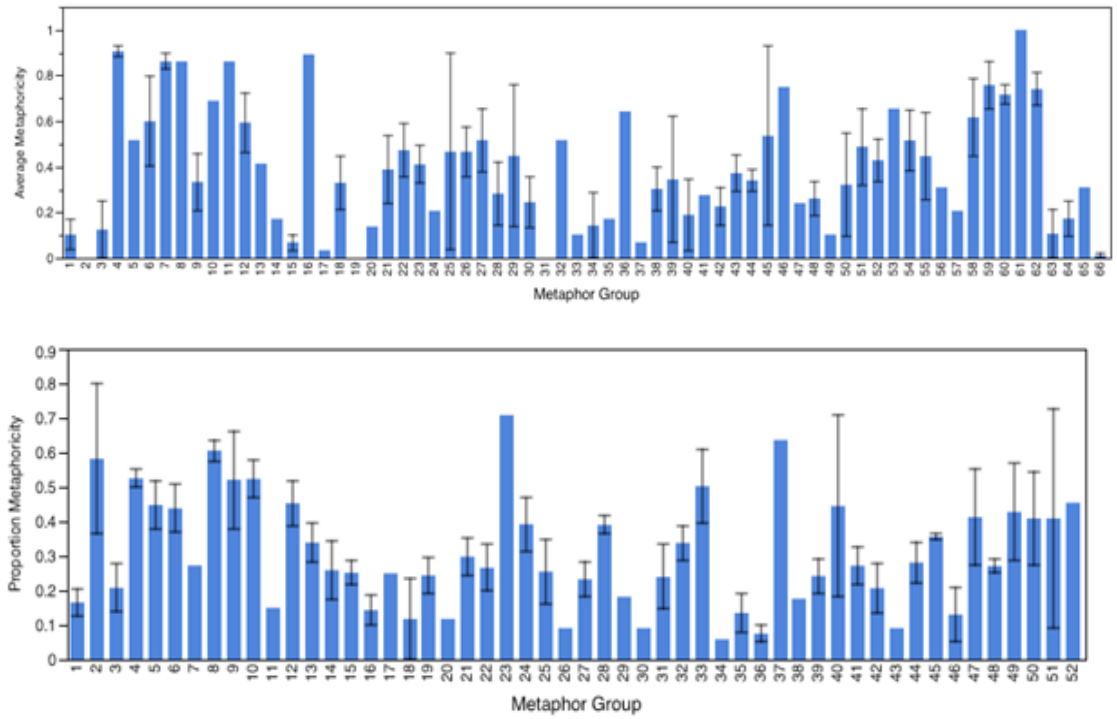

Figure 4: Metaphoricity of Conceptual Metaphors for English (top) and Spanish (bottom)

English Results

We restricted our subjects to those claiming to be native English speakers who had IP addresses within the U.S. and had 115 participants. The examples were grouped into 66 conceptual metaphors. The mean metaphoricity of instances was 0.41 (standard deviation=0.33). The mean metaphoricity of the conceptual metaphors (**Figure 4**), was 0.39 (SD=0.26). Although there was wide variance in metaphoricity across conceptual metaphors, it appears likely that most of the conceptual metaphors discovered by the system are correct: 65% of the conceptual metaphors had metaphoricity greater than 0.25, and 73% greater than 0.2. Given that many metaphors are conventional and difficult to detect in natural language (Lakoff and Johnson, 1980), it is possible that even in cases in which only a minority of subjects detected a metaphor, a metaphor nonetheless exists

Spanish Results

We restricted our subjects to those claiming to be native speakers of Mexican Spanish with IP addresses in the US (57) or Mexico (29). The instances were grouped into 52 conceptual metaphors. The mean metaphoricity of instances was 0.33 (SD=0.23) and for conceptual metaphors (**Figure 4**), 0.31 (SD=0.16). 60% of conceptual metaphors had metaphoricity greater than 0.25, and 73% greater than 0.2. That performance was only slightly lower than English is a positive indication of our method's cross-linguistic potential.

8 Discussion and Future Work

We observed a number of problems with our approach which provide avenues for future research.

8.1 Topics as Proxies of Primary Metaphor Concepts

Many of the metaphors missed by our system were instances of primary metaphor, especially those involving movement and spatial position. Our LDA approach is poorly suited to these because the source concepts are not well-characterized by word co-occurrence: words describing movement and spatial position do not have a strong tendency to co-occur with other such words, at least in Wikipedia. Augmenting our system with a separate

approach to primary metaphor would boost its performance significantly.

8.2 Topics as Proxies of Non-Primary Metaphor Concepts

We found that most of our potential source concepts did not correspond to any LDA topic. However, many of these, such as *wild west*, have fairly strong word co-occurrence patterns, so they plausibly could be found by a different topic modeling algorithm. There are two promising approaches here which could potentially be combined. The first is to use a hierarchical LDA algorithm (Blei et al, 2003b) to allow concepts to align to topics with varying degrees of granularity, from the very general (e.g. *war*) to the very specific (e.g. *wild west*). The second is to use constrained LDA approaches (Andrzejewski and Zhu, 2009; Hu et al., 2010) to attempt to force at least one topic to correspond to each of our seed concept lists.

A different approach would leave behind seed lists entirely. In our current approach, only about one third of the topics modeled by LDA are successfully aligned with a source concept from our hand-made list. However, some non-aligned LDA topics have properties similar to those that were chosen to represent source concepts. For instance, the topic whose highest ranked terms are [*institute, professor, engineering, degree*] is comprised of a set of semantically coherent and concrete terms, and could be assigned a reasonably accurate label such as *higher education*. If we were to choose LDA topics based on the terms' coherence and concreteness (and perhaps other relevant, measurable properties), then assign a label using a method such as that in Mei et al. (2007), we would be able to leverage more of the concepts in the LDA model. This would increase the recall of our system, and also reduce some of the confusion associated with incorrect labeling of concepts in linguistic and conceptual metaphors. Applying Labeled LDA, as in Ramage et al. (2009), would be a similar approach.

8.3 Confusion of Literal and Metaphorical Usage of Source Concepts

Another major problem was the confusion between literal and metaphorical usage of source terms. This is partly addressed by our document topics filter, but more sophisticated use of document context for this purpose would be helpful. A similar filter based on contexts across the test corpus might be useful.

8.4 Fixed Expressions

Some of our errors were due to frequent fixed phrases which included a word strongly associated with a source topic, like *Tea Party*. Minimum description length (MDL) phrase-finding or similar techniques could be used to filter these out. Initial experiments performed after the evaluations discussed above show promise in this regard. Using the MDL algorithm (Rissanen, 1978), we developed a list of likely multi-word expressions in the Wikipedia corpus. We then concatenated these phrases in the Wikipedia corpus before LDA modeling and in the test corpus before metaphor prediction. Though we did not have time to formally evaluate the results, a subjective analysis showed fewer of these fixed phrases appearing as indicators of metaphor (as words in A_s or A_T).

8.5 Difficulty of Annotation

A different method of presentation of metaphors to the subjects, for instance with annotations marking where in the sentence we believed metaphor to exist or with a suggestion of the source concept, may have improved agreement and perhaps the system's evaluation score.

8.6 Summary

We have presented a technique for linguistic and conceptual metaphor discovery that is cross-linguistically applicable and requires minimal linguistic resources. Our approach of looking for overlapping semantic concepts allows us to find metaphors of any syntactic structure. The framework of our metaphor discovery technique is flexible in its ability to incorporate a wide variety of source and target concepts. The only linguistic resources the system requires are a corpus of general-knowledge text adequate for topic modeling and a small set of seed word lists. We could improve our system by applying new research in automatic topic modeling, by creating new filters and scoring mechanisms to discriminate between literal and figurative word usages, and by creating training data to allow us to automatically set certain system parameters.

Acknowledgements

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number **W911NF-12-C0-0023**. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government."

References

David Andrzejewski and Xiaojin Zhu. 2009. Latent Dirichlet Allocation with Topic-in-Set Knowledge. In *Proceedings of NAACL Workshop on Semi-Supervised Learning for NLP*.

Collin Baker, Charles Fillmore, and John Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of COLING-ACL*.

Stephen Bethard, Vicky Tzuyin Lai and James H. Martin. 2009. Topic Model Analysis of Metaphor Frequency for Psycholinguistic Stimuli. . In *Proc. Of NAACL-HLT Workshop on Computational Approaches to Linguistic Creativity*.

Julia Birke and Anoop Sarkar. 2006. A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. In *Proceedings of EACL*.

David Blei, Thomas Griffiths, Michael Jordan, and Joshua Tenenbaum. 2003a. Hierarchical topic models and the nested Chinese restaurant process. In *Proceedings of NIPS*.

David Blei, Andrew Ng, and Michael Jordan. 2003b. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 2003(3):993–1022.

Dan Fass. 1991. met*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics*, 17(1):49–90.

Christine Fellbaum. 1998. WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA.

Matt Gegigan, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching Metaphors. In *Proceedings of the 3rd Workshop on Scalable Natural Language Understanding*.

Joseph E. Grady. 1998. Foundations of meaning: Primary metaphors and primary scenes. UMI.

Yuenin Hu, Jordan Boyd-Graber, and Brianna Satinoff. 2010. Interactive Topic Modeling. In *Proceedings of ACL*.

Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting Elusive Metaphors Using Lexical Resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*.

George Lakoff and Mark Johnson. 1980. Metaphors We Live By. University of Chicago.

James H. Martin. 1994. MetaBank: A knowledge-base of metaphoric language convention. *Computational Intelligence*, 10(2):134–149.

Zachary Mason. 2004. CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System. *Computational Linguistics*, 30(1):23–44.

Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu.

Qiaozhu Mei, Xuehua Shen, and Chengxiang Zhai. Automatic Labeling of Multinomial Topic Models. In *Proceedings of KDD '07*. 2007.

Wim Peters and Ivonne Peters. 2000. Lexicalised Systematic Polysemy in WordNet. In *Proceedings of LREC*.

Daniel Ramage, David Hall, Ramesh Nallapati and Christopher D. Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of EMNLP*.

Jorma Rissanen. Modeling by shortest data description. *Automatica* 14:465-471.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor Identification Using Noun and Verb Clustering. In *Proceedings of COLING*.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2012. Statistical Metaphor Processing. *Computational Linguistics*. Uncorrected proof.

Ekaterina Shutova. 2010. Models of metaphor in NLP. In *Proceedings of ACL*.

Joanna Thornborrow. 1993. Metaphors of security: a comparison of representation in defence discourse in post-cold-war France and Britain. *Discource & Society*, 4(1):99–119