Cross-linguistic automated detection of metaphors for poverty and cancer[1]

OANA DAVID

AND

TEENIE MATLOCK

Cognitive and Information Sciences, University of California, Merced

ABSTRACT

Conceptual metaphor research has benefited from advances in discourse analytic and corpus linguistic methodologies over the years, especially given recent developments with Natural Language Processing (NLP) technologies. Such technologies are now capable of identifying metaphoric expressions across large bodies of text. Here we focus on how one particular analytic tool, MetaNet, can be used to study everyday discourse about personal and social problems, in particular, poverty and cancer, by leveraging reusable networks of primary metaphors enhanced with specific metaphor subcases. We discuss the advantages of this approach in allowing us to gain valuable insights into cross-linguistic metaphor commonalities and variation. To demonstrate its utility, we analyze corpus data from English and Spanish.

1. Introduction

Metaphor is more than a rhetorical flourish. It is grounded in our everyday embodied and emotive experience, and drives much of the way we communicate (Gibbs, 1994; Johnson, 1987; Kövecses, 2015; Lakoff, 1987; Lakoff & Johnson, 1980, 1999). We use metaphor to structure our understanding of one domain in terms of another, for instance, time as space (Radden, 2011), anger as heat (Lakoff & Kövecses, 1987), desire as hunger (Gibbs, Lima, & Francozo, 2004), and politics as sports (Semino & Masci, 1996). Metaphor can have a dramatic impact on how we understand important social matters, such as crime (Thibodeau & Boroditsky, 2011, 2013), climate change (Flusberg, Matlock, & Thibodeau, 2017), and elections (Burnes, 2011).

Extensive research has examined metaphor across genres, especially in English. Manual methods, including qualitative analyses that use hand-annotation of texts, and corpus methods have been used to analyze when and how metaphor is used. The Metaphor Identification Procedure (Pragglejaz Group, 2007; Steen, 1999; Steen, Biernacka, Dorst, Kaal, López-Rodríguez, & Pasma, 2010a; Steen, Dorst, Herrmann, Kaal, & Krennmayr, 2010b), is extensively used in the broader metaphor community (e.g., Demjén, Semino, & Koller, 2016; Demmen et al., 2015; Steen, Dorst, Herrmann, Kaal, Krennmayr, & Pasma, 2010c). The second approach falls within more traditional corpus linguistic methodologies (Deignan, 2005; Lederer, 2013; Martin, 2006; Philip, 2004; Stefanowitsch & Gries, 2006), which use concordances, frequency counts, keyword analysis, and collocation patterns to identify potentially metaphoric uses of target words.

With more and more attention on scalability to larger texts, a third approach is emerging from computational linguistics. It focuses on devising automated means of metaphor identification across larger or more diverse corpora. Metaphor is a burgeoning area in the field of computational linguistics, evident from the addition of a figurative language and metaphor processing workshop to the annual meeting of the North American Chapter of the Association for Computational Linguistics (NAACL <www.naacl.org>). While specific methodologies and goals vary, most NLP approaches aim for improvements in recall and precision mechanisms for automated linguistic metaphor identification based on internally defined retrieval definitions. The automation is usually exercised over unrestricted texts, texts that are sufficiently general to train a system to identify all types of metaphoric expressions. Some studies use statistical cluster methods

468

(Birke & Sarkar, 2006; Shutova & Sun, 2013; Shutova, Teufel & Korhonen, 2012) for metaphor identification. Gutiérrez, Shutova, Marghetis, and Bergen (2016) employ compositional distributional semantic vector space models that operate on the composition of lexical representations across large corpora. CorMet (Mason, 2004) uses the selectional preference of verbs, and clusters of nodes derived from WordNet senses. There are many other methods explored in computational metaphor detection: neural nets (Do Dinh & Gurevych, 2016), maximum entropy classification combined with hand-annotation of metaphoricity (Gedigian, Bryant, Naryanan, & Ciric, 2006), selectional preferences of lexical items (Wilks, 1975, 1978), knowledge-representation models (Martin, 1988, 1994), and word sense disambiguation-based approaches (Krishnakumaran & Zhu, 2007), among others. Dunn (2013a, 2013b) describe four systems of metaphor identification and their comparison in performance, while Shutova (2010) and Neuman et al. (2013) provide overviews of the current state of the art in metaphor NLP.

The above are just a few of the approaches and technologies available in this quickly expanding subfield of computational linguistics. Such automation approaches are useful because they are agnostic to specific research needs as well as specific textual genres, and are capable of yielding high scores for recall and precision in automated metaphor identification due to their rigorous implementation of lexical and semantic resources (e.g., SOMO ontology in Dunn, 2013a, and WordNet in Lönneker, 2003). However, they may not be ideal when a metaphor researcher has specific questions about the functions of metaphor in a particular language context or cognitive and social domain, or seeks a bird's-eye view of metaphor distribution within that domain.

Another area in which existing computer-aided metaphor research could be enhanced is in cross-linguistic studies. With some recent exceptions (Gordon, Hobbs, May, Morbini, & Vista, 2015; Levin et al., 2014; MacWhinney & Fromm, 2014; Mohler, Tomlinson, & Rink, 2015; Tsvetkov, Boytsov, Gershman, Nyberg, & Dyer, 2014),[2] metaphor NLP pipelines are created for detecting metaphor in one language, usually English. This constraint results from the nature of the lexical resources on which they are trained, as English is a language for which lexical resources, and NLP resources in general, are more developed relative to other languages.

---

[2] Not uncoincidentally, these works are all the results of sister projects to the MetaNet system described here. All are the result of a grant from the Intelligence Advanced Research Projects Activity (IARPA), which required cross-lingual coverage as part of the evaluation for success of each project. For details on the history and outcome of the larger project, see Sweetser et al. (in press).

MetaNet, the system of interest here, can search for cross-linguistic metaphor distribution in texts and narrow down the search to a particular domain of interest. It seeks to form a more symbiotic relationship between powerful computational models on the one hand, and on the other, methods of studying metaphor that have as their goal an understanding of cognitive and social realities. It is not intended to be a general NLP system like the ones described above, rather, a tool that helps metaphor analysts carry out large-scale studies, for instance, by narrowing the search according to target domain and providing cross-linguistic metaphor distribution. The ideal user is a researcher seeking an initial representation of metaphor distribution in specific domains (for instance, as shown here, in domains of social concern such as cancer and poverty) over large texts, and possibly over multilingual corpora. The distribution obtained would not paint an exhaustive picture of metaphor in those texts, as the tool cannot detect every possible metaphor, but it would be representative enough to enable the researcher to pursue particular hypotheses, or to focus on particular subdomains or metaphor families and text genres for further probing. Sweetser, David, and Stickles (in press) provide a thorough comparison of MetaNet with other automated metaphor detection systems, including those listed above.

Each automated identification run begins with a specific question issued by a researcher, either about how a target domain is talked about metaphorically (via target domain search, e.g., cancer), or how a source domain is used across several target domains (via a source domain search, e.g., Forward Motion or Physical Combat). Further, repeated iterations of MetaNet-aided metaphor extraction help augment its knowledge base, such that an increasing number of metaphors across an increasing number of domains (and languages) are accumulated and reused in future iterations. Gold standard annotations were used at multiple junctures to evaluate the system for precision and recall scores, and to reduce the detection of false positives over time (see Hong, 2016). However, since the system is only trained to detect metaphors in a finite set of grammatical patterns (Dodge, Hong, & Stickles, 2015; Stickles, David, Dodge, & Hong, 2016) and for a finite set of metaphor families, some occurrences of domain-relevant metaphor may not be detected. For instance, *give cancer the boot* would not be detected because the ditransitive construction is not yet represented in the system. Nevertheless, this limitation does not encumber the system's ability to give a good idea of the broad distribution of metaphors in the domain of interest by producing results on a large scale.

Here, we briefly summarize the architecture of metaphor detection automation in MetaNet. We focus on two implementations of this system for metaphor discovery. Our main interests are to showcase some results produced with this system, and to delve into some insights about cross-domain similarities, specifically in the domains of cancer and poverty. Prior work offers detailed

discussion of the MetaNet architecture, so there is no need to focus on it here. Hong (2016) discusses the MetaNet pipeline, including how error (false positives and negatives) is handled, how metaphoricity scoring is determined, how gold standard annotations are used to train the system, and how precision and recall (f-score) are measured. The ontological structures and hierarchically organized knowledge base forming the metaphor and frame repository are detailed in Stickles et al. (2016). The latter provides a discussion of similarities and differences with other frame-representational lexicographic resources such as FrameNet (Ruppenhofer, Ellsworth, Petruck, Baker, & Scheffczyk, 2016) and WordNet (Fellbaum, 1998), as well as the role of grammatical constructions in determining the relationship between lexical items and source-to-target domain mappings. This paper also outlines taxonomies of frame-to-frame and metaphor-to-metaphor relations, and discusses how decisions were made on what counts as a conceptual metaphor, and how conceptual metaphors are assigned to linguistic instantiations. Additional discussion on the lexical and frame resources seeding MetaNet's frame repository is found in Dodge et al. (2015), which includes examples of how recall and precision are evaluated, with special attention to poverty. David, Lakoff, and Stickles (2016) provide a detailed sketch of how primary metaphor networks are designed to find data in the domain of gun rights, showing how existing primary structures are augmented (with new frames, new specific metaphors, and new lexical items) to enhance coverage. Similarly, David (2017) illustrates MetaNet's metaphor network expandability with data from the target domain of democracy.

2. Theoretical grounding

Primary metaphor (Grady, 1997) figures prominently in the network of inter-related metaphors in the metaphor repository. They serve as the basis for more specific metaphors. Primary metaphors are schematic, embodied, and likely to be universal (Kövecses, 2005; Lakoff, 2012). They are formed early in cognitive development, and result from predictable simultaneous co-experiences of two domains during interactions with entities, forces, and people in the world (Grady, 1997; Johnson, 1987). Primary metaphors, such as STATES ARE LOCATIONS, PURPOSES ARE DESTINATIONS, MORE IS UP/LESS IS DOWN, and COMMUNICATION IS OBJECT EXCHANGE are an unchanging part of this core repository, while additional metaphors may be added as new linguistic metaphors about specific target domains are discovered via empirical methods. For instance, a phrase like *tackle poverty* instantiates the primary metaphor DIFFICULTY IN ACTION IS PHYSICAL COMBAT, but it also fulfills a more specific metaphor, DEALING WITH POVERTY IS PHYSICAL COMBAT. The latter is a more specific subcase of

471

the former and a closer match for the identified linguistic metaphor. The verb *tackle* evokes the source domain of Physical Combat, as do many other verbs such as *combat*, *fight*, *attack*, and *defeat*. By associating multiple lexical items with particular source domains, and setting those domains as the source domains of metaphors, the system is trained to find tokens on the basis of one type (e.g., *tackle poverty*) such that it finds multiple other types (using different verbs) and tokens while still fulfilling the same source domain. The association of multiple lexical units to a single frame is a procedure inherited from other frame-based approaches to lexicography and semantic computation, such as FrameNet (Ruppenhofer et al., 2016), but takes one step further by associating frames with conceptual metaphors (frame-to-frame mappings).

Semantic frames are central to defining metaphor in MetaNet. A semantic frame is a knowledge schema whereby "a word's meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning" (Fillmore & Atkins, 1992, pp. 76–77). It is akin to scheme, script, or frame in psychology (Barsalou, 1982), and emerged as a part of FrameNet and other computational and lexicographic resources (Fillmore, Johnson, & Petruck, 2003; Ruppenhofer et al., 2016). The utility of frames for modeling metaphoric mappings has been considered (Sullivan, 2006, 2016), especially at the intersection of conceptual metaphors and lexico-grammatical structures. MetaNet models metaphor as frame-to-frame mappings, and therefore uses frames, frame elements, and lexeme-to-frame evoking patterns. It also uses grammatical constructions to determine how mapping in metaphoric words occurs, given their syntactic environment (David, 2016; Stickles et al., 2016).

3. Corpora and semantic resources

Of the two case studies presented here, one compares English and Spanish metaphors for poverty in the Gigaword corpora (Mendonça, Jaquette, Graff, & DiPersio, 2011; Parker, Graff, Kong, Chen, & Maeda, 2011), and the other compares cancer metaphors in English in two corpora, a general corpus (the GLoWbE corpus: Davies, 2013; Davies & Fuchs, 2015) and a specialized corpus compiled by the MetaNet team, consisting only of cancer blog texts. The corpora are summarized in Table 1.

In the first study, we test the system's efficacy in a cross-linguistic comparison within one domain (poverty), and in the second, we perform a cross-corpus comparison in another domain (cancer). The cross-linguistic (poverty) comparison is between two corpora of the same size, and therefore the raw counts are comparable. The Gigaword corpora in English and Spanish were chosen because of their availability in both languages, their large sizes,

472

TABLE 1. *Summary of corpora used*

	Cancer	Poverty
English	GloWbE: 1.3 million tokens (blogs only) (1.9 billion total)	English Gigaword (5th edition): ~4 billion words Cancer blog and forum corpus: 703,000 words
Spanish	n/a	Spanish Gigaword (3rd edition): ~4 billion words

and their representativeness in the domain of poverty, a topic that often appears in the type of newswire data in these corpora.

The cross-corpus (cancer) comparison is between a very large and a much smaller corpus, and therefore the results are presented as normalized per 100,000 tokens. Since the custom cancer corpus was collected from blog and forum entries, thus constituting a coherent genre, the GloWbE English corpus was selected for its ability to be filtered by the blog genre, providing a comparable subcorpus. The GloWbE corpus was mined only for a subset of data tagged as blogs because it is quite large, and because genres other than blogs are more general and not necessarily venues for writing about cancer.

The specialized cancer corpus was compiled to have a high concentration of cancer discourse, collected from online cancer discussion boards (the domains scraped are <https://community.breastcancer.org> and <www.cancerforums.net>). Genre-based comparisons in corpus linguistics help determine whether the observed trends are a feature of the language in general, or of a particular sampled speech community in particular. The discussion boards mined provide cancer patients, loved ones, and doctors with a venue for sharing information about treatment, venting frustrations with side effects and relapses, sharing good news, asking informational questions, and creating a sense of community where sensitive topics can be discussed without fear of judgment.

The resources discussed here are available online. The MetaNet source code is available in a GitHub repository (https://github.com/TheMetaNetProject), and a publicly searchable MetaNet Wiki, containing all the metaphors and frames (English only), is located at <https://metaphor.icsi.berkeley.edu/pub/en/>.

4. Cross-domain comparability in metaphors for social issues

MetaNet is a large-scale automated metaphor identification system that can be used to better understand non-literal language use in two different domains, poverty and cancer. Before discussing our analysis, we consider why it is useful

473

to study these domains together, and why we might expect them to have anything in common.

Both cancer and poverty are social problems, and thus, the metaphors used to communicate about them often lead to negative inferences about their effects on individuals and on society. For instance, dealing with poverty and cancer is frequently seen as a physical struggle (*combat poverty*, *fight against cancer*), and as movement (*get out of poverty*, *my cancer journey*). Studying them together gives a better sense of how high-level primary metaphors (such as DIFFICULTIES IN ACTION ARE PHYSICAL STRUGGLES WITH AN OPPONENT and PROGRESS IS FORWARD MOTION) behave across multiple discourse domains and genres. Rather than arguing that for two separate metaphors, CANCER IS WAR and POVERTY IS WAR, we can state a generalized metaphor such as SOCIAL/PERSONAL DIFFICULTIES ARE PHYSICAL COMBAT. In so doing, we can use the same concrete domain inferences to reason about multiple target domains, possibly opening the door to others in the future (e.g., terrorism, drugs, addiction).

That poverty should be metaphorically construed as a personal and social struggle is unsurprising – poverty is indeed a major problem. How is metaphor used in discourse about poverty, and how does it shape policy, ultimately? First, in descriptions of poverty in the media we often observe a preponderance of violence metaphors. President Lyndon B. Johnson's famous 'war on poverty' slogan resulted in a national-wide effort to target many aspects of poverty. It led to the creation of long-lasting national policies (see Cahn & Cahn, 1964). In this case, poverty is framed as the enemy, and the resisting protagonists are individuals or agencies who have to fight poverty.

Recent work on metaphors for poverty using the MetaNet system (Dodge, 2016) also shows how it is often framed as some type of harm (including disease) or movement, usually to and from a physically low location, e.g., *fall into poverty*, *get out of poverty*. As we illustrate here, these metaphors appear in both English and Spanish, with greater or lesser degrees of violence evoked by the lexical items used. Some metaphors, as in (1) and (2), suggest general physical antagonism, while others, like the Spanish and English examples in (3) and (4), are more strongly militaristic and evoke either weaponized combat or hand-to-hand combat.[3]

---

[3] The example sentences throughout are results from the MetaNet automated metaphor identification process. Each example will follow the naming convention: *source corpus name_ID number*. The source corpora are: ENGW: English Gigaword, ESGW: Spanish Gigaword, GLO: GloWbE blogs corpus, CAN_SPEC: specialized cancer corpus compiled for this work. The ID number is a unique identifier within the system for each linguistic metaphor automatically detected.

(1) The agenda for an upcoming gathering of conservative Anglican clerics includes discussions about dialogue with Islam and *fighting poverty*. (ENGW_72353)

(2) Feminists also used the occasion to draw attention to the estimated 100,000 Filipinas *forced by poverty* to work as prostitutes, some of them exported to Japan and other countries. (ENGW_40)

(3) La *miseria invade* a los libaneses desplazados de Trípoli.
'*Poverty invades* Lebanese displaced out of Tripoli.'
(ESGW_501479958)

(4) A job is the best *weapon against poverty*. (ENGW_16010)

Nevertheless, though metaphors of fighting, war, and violence are common, inspection of the data reveal that other metaphors are used as well. In (5) and (6), for instance, poverty is construed as a low location into which one can sink, and from which one hopes to emerge.

(5) … ha enfrentado graves conflictos políticos, militares y sociales, en un país *sumido cada vez más en la pobreza*. (ESGW_502037711)
'… (he) has faced serious political conflict, both military and social, in a country *immersed (deeper and deeper) in poverty*.'

(6) He also estimated massive aid from outside was needed to help the African nations *emerge from poverty* and support political and economic reforms. (ENGW_145)

Another common source domain for construal of poverty is disease, one that can infect individuals and societies, and one that must be addressed with metaphorical treatments and medicine, usually in the form of strong socioeconomic reform. As shown in (7), speakers often mix metaphors. In mixing war metaphors with disease metaphors they construe the social condition simultaneously as one that needs to be combatted and one that is a pathology.

(7) We must *combat* such social *pathologies* as widespread poverty, the breakdown of family life, crime, alcohol and drug abuse. (ENGW_56)

Indeed, framing crime as a disease or virus can cause people to think about systemic, reform-oriented solutions, and framing it as a beast or monster can cause them to believe harsher punishments are needed (Thibodeau & Boroditsky, 2013). Since the corpus work presented here reveals that poverty is also widely metaphorically framed as a disease, this could inspire new empirical work on whether similar patterns of reasoning will emerge with respect to this socioeconomic issue.

Like poverty, we have reason to believe that cancer, as a personal and a social problem, is metaphorically construed using two prominent metaphors, CANCER IS A WAR (or PHYSICAL COMBAT) and CANCER IS A JOURNEY

475

(Demmen et al., 2015, Semino et al., 2015). Indeed, in a similar spirit to President Johnson, President Nixon declared a 'war on cancer' only seven years later (National Cancer Institute, 1971), and many other 'wars' on many other perceived social ills have been waged since then, from the war on drugs to the war on terror (Elwood, 1995).

As is clear from the following excerpt from our specialized cancer blog corpus, war language is not unique to leaders talking about policy goals as a nation. It also appears in individuals' discussions of their subjective, personal experience with the disease, as in (8).

(8) I describe my year of cancer as *sleeping with an enemy*, *unrelenting and omnipresent*. I couldn't shake *its presence*, even in those brief moments of relaxation. […] To the world, I presented as *unsinkable*. This is what my dear family and friends required. They needed their Amy back, the Amy who always *marches on*, through the best and the worst. Probably, I needed it, too. Everyone felt better telling me *how strong I am* and how *I will beat it*, but there, in the dark, I was terrified and helpless. […] Each time I went for chemo, I felt like *a lamb going in for slaughter*. I tried to visualize *tiny resistance fighters* living in my breast, *my own Polish forest, beating away the Nazis* inside of me. (Small-McKinney, 2014)

In this excerpt, the writer introduces cancer as an enemy, a persistent undesirable, evil companion to fight. She refers to expectations for her to be a soldier in the battle, one who will be strong and continue marching on despite feelings of fear and helplessness. She refers to her immune system as small soldiers who fight the cancer inside her breast, which she depicts as a Polish forest, referring back to a WWII location that saw heavy losses. Invoking military history allows her to vivify her experience with cancer, including her feelings of helplessness. Often, as the *lamb for the slaughter* idiom above shows, the metaphor shifts to other, non-militaristic yet still violent images, sustaining the struggle dynamics introduced at the beginning.

As (8) illustrates, cancer treatment is often discussed in terms of battles in the popular media, among friends and family members, and in doctor–patient interactions (Magaña & Matlock, 2018; Olweny, 1997; Semino et al., 2015). Common in such discourse is language such as *fighting the disease*, *knocking the cancer down*, *army of oncologists*, and *winning the battle*. Indeed, war language is predominant in metaphors for all types of diseases (Casarett et al., 2010; Weiss, 1997), as well as for dealing with pain in general (Semino, 2010; Stewart, 2014).

There are other types of metaphors for both poverty and cancer as well as for other social issues. Common metaphoric source domains include measurable objects with size and area (*cancer/poverty grows*, *cancer/poverty spreads*), vision-related domains (*live in the shadow of poverty/cancer*), and living entities

476

(*the cancer pest*, *the jaws of poverty*). Nevertheless, while some diversity exists, experimental and corpus research has shown that social, psychological, and other non-tangible phenomena perceived as detrimental in some way are often metaphorically characterized using one or two of a very limited set of source domains. For instance, experiments show that people can naturally be nudged towards policy change decisions when crime is construed as either a beast or a disease (Thibodeau & Boroditsky, 2011), and the conceptualization of climate change has similar salient dual construal patterns, i.e., dealing with climate change is seen either as war or as a race (Flusberg et al., 2017).

To deduce, using traditional corpus methods, what the metaphors for poverty, climate change, or other domains might be, we would have to perform individual studies on each domain and compare findings. This would be challenging from a standpoint of methodological comparability, as each study could be carried out according to different criteria, using different linguistic triggers, and based on varying decisions regarding determination of metaphoricity (see Gibbs, 2015, and Semino, Heywood, & Short, 2004, for commentary on problems inherent in metaphor analysis).

We present a multidomain, multilingual study using the automated metaphor identification method provided by MetaNet. In the next section we detail some of the inner workings of this system, and illustrate how it goes about finding metaphor in texts.

5. The MetaNet architecture and procedure

MetaNet consists of two parts: the metaphor repository and the automated metaphor identification processor. The repository is a network of hierarchically organized metaphors. It contains relationships between lexical items and the source- and target-domain frames linked to those metaphors. For instance, the metaphor DEALING WITH CANCER IS PHYSICAL COMBAT is a subcase of the metaphor DIFFICULTIES IN ACTION ARE PHYSICAL COMBAT. The lexical items and phrases *cancer*, *cancerous*, *cancer treatment*, etc. are associated with the Cancer frame, the target-domain frame of the first metaphor. The lexical items *fight*, *attack*, *punch*, etc. are associated with the Physical Combat source domain frame of both metaphors. All of these entities and their relationships are stored in the repository, which later acts as a knowledge base for the automated metaphor identification processor to perform its function.

The scripts in the automated metaphor identification processor use information from the repository to crawl over corpora and find metaphoric expressions within a limited set of domains (Hong, 2016). The domains currently covered by MetaNet are include poverty, gun control, democracy, taxation, governance, and bureaucracy and other domains of social concern.

Because these problems receive much coverage in the media, there is ample opportunity to scrape data from relevant news reports. Together, the two parts yield a third component, the annotation database, which stores the automatically identified metaphoric sentences, assigning them internal ID numbers (reported along with example sentences in the current work), and provides information about what source domain word (e.g., *crushing*) and what target domain word (e.g., *poverty*) were involved.

The architecture is set up so the source and target domain frames of metaphors are arranged in an inheritance lattice relative to each other, in which specific frames are subcases of more general ones. For instance, as shown in Figure 1, Poverty is a specific type of Social Problem, and Disease, a more specific type of Physical Affliction. Metaphors are also hierarchically organized. For instance, POVERTY IS A DISEASE is a subcase of SOCIAL PROBLEMS ARE PHYSICAL AFFLICTIONS.

When a metaphoric collocation is encountered in text, e.g., *cure poverty*, at least one of two metaphors is automatically realized. The metaphor POVERTY IS A DISEASE, if present in the metaphor repository, would be recognized via link number 2 in Figure 1 because Poverty and Disease would be directly linked to that metaphor as target and source domain frames, respectively. However, if such a specific metaphor is unavailable, we at least get the more general SOCIAL PROBLEMS ARE PHYSICAL AFFLICTIONS as a candidate metaphor, via the links labeled 1 in the diagram. This way, because Poverty is a subcase of another higher-level frame (which itself is the target domain frame of a higher-level metaphor), we achieve accurate metaphor detection from the lexical inputs of *cure* and *poverty* at varying levels of specificity.

This mechanism enables queriers to use few metaphor networks to identify many metaphoric expressions across diverse texts. The system is multilingual, and thus able to dispatch a core set of conceptual metaphors over multiple languages (e.g., English and Spanish). Data presented below were identified as linguistic metaphors with metaphoricity scores of 0.7 or higher.
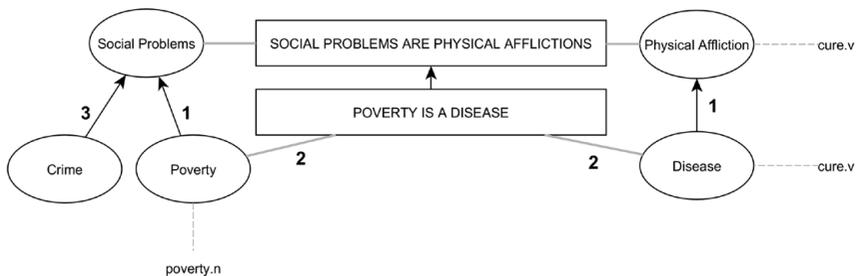

Fig. 1. Metaphor inheritance diagram for POVERTY IS A DISEASE with relations to lexical items.

478

Each automatically detected linguistic metaphor receives a metaphoricity score from 0 to 1, based on the best path through the metaphor and frame inheritance networks triggered by the lexical items.[4] Therefore, when literal uses of source domain-evoking terms are picked up, they are filtered out. This reduces the occurrence of false positives.

The methodology presented here differs from existing corpus-based approaches to metaphor (e.g., Stefanowitsch, 2006) in one crucial way: lexical items do not directly map to metaphors, and are not themselves taken as metaphoric for purposes of discovery. Instead, lexical items are associated with frames, which in turn are associated with metaphors; the metaphoricity of a phrase is determined by a cascade exploiting frame inheritance networks, metaphor inheritance networks, and frame-to-metaphor relationships, while mediating via the grammatical constructions in which the candidate expression appears (Dodge et al., 2015). The same lexical item, e.g., *cancer*, can be metaphoric in one sense (*poverty is a cancer in our society*) and literal in another (*a tough battle with cancer*), and the system can determine which is which. This method can be expanded to additional languages, especially given that lexical items are associated with existing frame and metaphor structures. Though no substitute for in-depth qualitative analysis that seeks a broad set of potentially culturally specific linguistic instantiations of these and other conceptual metaphors, this system can paint an initial picture of multilanguage distributions of metaphors known to be dominant in the domains of interest, especially when crawling over very large corpora in sizes (not possible with hand-annotation).

6. Automated metaphor identification results

6.1. POVERTY IN CROSS-LINGUISTIC COMPARISON

A survey of metaphor detection results for metaphor identification in the poverty domain from the Spanish and English Gigaword corpora reveals an abundance of metaphor around target domain expressions such as *poverty*, *destitution*, *impoverished*, *impoverishment*, *indigence*, *underprivileged*, *the 47 percent*, and *the 53 percent*, with similar expressions in Spanish. Table 2 shows the target domain-evoking lemmas, and the number of results of linguistic metaphors detected in the English and Spanish Gigaword corpora. These numbers illustrate that metaphoric language gravitates primarily around one or two key target domain terms.

Each target-evoking word and phrase is associated with frames that are associated with target domain frame slots of one or more metaphors

---

[4] For a precise description, see Hong (2016, Section 2.4) and Dodge et al. (2015, Section 3.2).

479

TABLE 2. *Raw frequencies of metaphoric poverty-related expressions occurring with specific target lemmas. Metaphoric senses are most common for* poverty *and* pobreza, *with some occurrences across other lemmas*

	AFP	APW	XIN	CNA	LTW	NYT	WPB	Total
destitution	43	35	37	0	19	73	2	209
impoverished	395	457	54	4	122	307	4	1,343
impoverishment	26	21	16	0	13	45	0	121
indigence	0	1	3	0	1	6	1	12
poverty	10,639	11,748	14,510	153	3,016	9,929	186	50,181
poverty line	878	561	1,066	21	180	493	6	3,205
poverty rate	94	178	231	0	122	427	16	1,068
the 47 percent	4	2	0	1	2	13	0	22
the 53 percent	13	3	2	0	1	11	0	30
underprivileged	51	58	25	15	15	61	215	440
Total	**12,143**	**13,064**	**15,944**	**194**	**3,491**	**11,365**	**430**	**56,416**
empobrecer	24	20	6	–	–	–	–	50
empobrecimiento	67	12	13	–	–	–	–	92
indigencia	156	60	110	–	–	–	–	326
indigente	240	108	54	–	–	–	–	402
mendicidad	30	10	10	–	–	–	–	50
miseria	1,014	351	172	–	–	–	–	1,537
penuria	393	148	22	–	–	–	–	563
pobre	8,946	4,276	2,713	–	–	–	–	15,935
pobreza	10,754	5,471	5,430	–	–	–	–	21,655
pobreza extrema	606	158	240	–	–	–	–	1,004
privación	208	90	39	–	–	–	–	337
Total	**22,438**	**10,704**	**8,809**	**–**	**–**	**–**	**–**	**42,304**

NOTES: AFP: Agence France-Presse, English Service; APW: Associated Press Worldstream, English Service; XIN: Xinhua News Agency, English Service; CNA: Central News Agency of Taiwan, English Service; LTW: Los Angeles Times/Washington Post Newswire Service; NYT: New York Times Newswire Service; WPB: Washington Post/Bloomberg Newswire Service.

(e.g., *poverty* (n.) → Poverty frame → DEALING WITH POVERTY IS PHYSICAL COMBAT). For this sample search, the target-evoking words and frames are fixed to isolate only those metaphors pertaining to poverty, as many other metaphors are in the system (e.g., communication metaphors, emotion metaphors) and they may be using the combat source domain as well. The metaphors are objects the system uses to create a link between the target and source domain frames, with the latter bringing a set of associated lexical items. For the poverty domain, the stored metaphors are POVERTY IS A CRIME, POVERTY IS A DISEASE, POVERTY IS A FIERCE CREATURE, POVERTY IS A MOTION IMPEDIMENT, POVERTY IS A PLANT, POVERTY IS PHYSICAL HARM, ADDRESSING POVERTY IS TREATING AN ILLNESS, ADDRESSING POVERTY IS WAGING WAR, AMOUNT OF POVERTY IS SIZE, and AMOUNT OF POVERTY IS SIZE OF GEOGRAPHIC FEATURE.

480

These metaphors are nested in more complex hierarchical networks of metaphor, such that, for instance, if the phrase *mushrooming poverty rate* is detected, it simultaneously counts as AMOUNT OF POVERTY IS SIZE (specific) and ABSTRACT QUANTITY IS AN OBJECT WITH MEASURABLE SIZE (general). The latter has further entailments, such as REDUCTION IN ABSTRACT QUANTITY IS DECREASING IN SIZE (e.g., *shrink poverty*). This entailment is not present for the poverty-specific subcase metaphor, but the latter can still benefit from it in the metaphor identification process, and a phrase such as *shrinking poverty* now can be detected as well. In short, while only a handful of metaphors are dedicated to poverty in the system, many more metaphors about poverty not limited to that list can be detected by virtue of the hierarchical way in which the metaphors are structured. The rich set of lexical items associated with source domain frames yields this amplified detection effect.

Once the set of metaphoric expressions sought are narrowed down to ones in which the items in Table 2 appear, it is time to detect any source domain language. Lexical items are grouped into frames, such that *mushroom*, *expand*, and *balloon* all evoke the frame Increase In Size.[5] In Table 3, the source domains of metaphors for poverty, categorized by subgroups and groups of metaphoric source domain frames, are reported with their normalized frequencies (NF) per 1,000 extraction results, for comparability between the two languages and example lexical unit (LU) types encountered in each language. The lexical units are the source-domain frame words and phrases that evoke the conceptual metaphors.

Table 3 captures important information about metaphor in these languages. At a high level, metaphors can be grouped into three broad categories that have something in common semantically among the source domain frames – Violence/ Harm, Location/Motion, and Properties of Objects and Entities. In Table 3, these groups are sorted from most to least amount of variation between English and Spanish in terms of the absolute difference (Abs. Dif.) between normalized frequencies in the two languages, per 1,000 results. These three broad categories were not instrumental in deriving the results, but are merely groupings provided for expository purposes so as to render the data easier to view.[6]

---

[5] The procedure adopted for the assignment of Lexical Unites to frames in MetaNet is similar to that observed for FrameNet, per Ruppenhofer et al. (2016). See Stickles et al. (2016) for a detailed discussion of the decision-making process.

[6] Frames in the system are defined at a more fine-grained level, and thus are too numerous to report individually. For this reason, the subgrouping and grouping is useful in order to report larger trends. For instance, the Crime subgroup consists of the frames Arresting (*arrest*, *catch*), Legal process (*condemn*, *sentence*, *verdict*), and Crime scene (*victim*, *hold-up*, *criminal*, *police*). With respect to metaphors about poverty, not all of these will necessarily occur in natural speech, therefore they are taken as a whole, to see which lexical items end up appearing in metaphoric collocations, e.g., *condemned to a life of poverty*.

481

TABLE 3. *Results for Poverty metaphors in English and Spanish from the Gigaword corpora (quartile ranges: 33–130, 10–32, 2.5–9, 0–2.5). NF: normalized frequency; LU: Lexical Unit; Abs. Dif.: absolute difference*

	English NF	English example source LUs	Spanish NF	Spanish example source LUs	Abs. Dif.
Violence/Harm					
Physical struggle	126.7	*tackle, fight*	52.0	*lucha, pegar*	74.7
War	43.4	*combat, battle*	117.7	*invadir, arma*	74.3
Disease	197.7	*afflict, plague*	131.4	*enfermo, crónico*	66.3
Controlling entity	14.9	*grip, clutch*	2.5	*liberar, agarrar*	12.4
Burden	16.1	*ease, burden*	5.5	*pesar, carga*	10.5
Crime	4.5	*condemn, sentence*	15.0	*condenar, castigar*	10.5
Physical harm	31.7	*suffer, hit, threat*	24.1	*sufrir, lastimar, doler*	7.6
Destruction	8.2	*erode, deteriorate*	1.3	*vorágine, destruir*	6.9
Physical competition	6.9	*challenge, win, lose*	8.1	*ganar, perder*	1.2
Fire	0.8	*spark, flame*	0.9	*infierno, quemar, llama*	0.1
Darkness	0.5	*shadow, cloud, darken*	0.5	*sombra, negro*	0.1
Tyranny	0.6	*bully, tyranny*	–	*N/A*	–
Theft	0.3	*rob, steal, pillage*	–	*N/A*	–
Total:	**452.2**		**359.1**		
Object/entity properties					
Growth in size	7.3	*expand, swell*	68.4	*engrosar, aumentar*	61.1
Reduction in size	20.0	*cut, slash, diminish*	33.6	*pequeño, estrechar*	13.6
Game	4.8	*deal, trump, gamble*	13.7	*torneo, partido*	8.9
Building	19.3	*fortress, dismantle*	27.6	*edificar, erigir*	8.3
Plants	17.8	*yield, sow, seed*	9.8	*cultivar, sembrar*	8.0
Machine	1.3	*fix, tool, engine*	3.8	*operar, herramienta*	2.4
Observed entity	0.2	*blind, vision*	9.5	*reconocer, vislumbrar*	9.3
Liquid	0.3	*ocean, flood*	6.5	*mar, flujo, liquidar*	6.2
Animal	3.6	*breed, jaw, claw*	6.4	*tentáculo, animal*	2.9
Monster	1.4	*evil, haunt, menace*	4.2	*bestia, amenaza*	2.7
Eating	1.8	*biting, bitter*	0.6	*consumir, devorar*	1.2
Pest	1.0	*blight, teem, infest*	–	*N/A*	–
Total:	**78.8**		**184.1**		
Location/Motion					
Level	61.3	*level, layer, tier, rung*	115.6	*nivel, línea*	54.3
Upward movement	57.0	*rise out of, lift*	4.9	*escalar, subir, arrancar*	52.1
Downward movement	77.3	*decline, plunge*	28.6	*bajar, caer, hundir*	48.6
Forward movement	35.6	*launch, crawl*	66.8	*marcha, pasar, avance*	31.2
Enablements to motion	6.7	*gateway, door*	21.6	*entrada, umbral*	14.9
Impediments to motion	26.7	*obstacle, (road)block*	12.7	*bloque, estancar*	14.0
Be located	141.8	*live in, linger, wallow*	152.7	*estar en, ubicar*	10.9
Path	9.8	*way out, route, ladder*	0.7	*camino, rastro*	9.1
Gap	3.5	*gap, divide, gulf*	2.1	*brecha, separación*	1.5
Confinement	46.6	*prison, trap, shackle*	47.8	*ligar, atrapar*	1.2
Maze	2.1	*web, entangle, maze*	2.5	*círculo, laberinto*	0.4
Backward movement	0.7	*fall back, recede*	1.0	*retiro, retirar*	0.3
Total:	**469.0**		**457.0**		

Second, the data are presented in cells according to quantiles of normalized frequencies (represented as varying cell shadings), and calculated based on log values, which are not reported. This ensures that ranges of normalized frequencies categorizable are presented in a comparable way, as the scales vary greatly from highest-frequency (in the hundreds, darker gray) to the lowest-frequency items (often below 4 occurrences per 1,000 results, unshaded). The main clusters are around the Location/Motion and Violence/Harm in both languages. For Violence/Harm, English is dominated by metaphors that construe poverty as physical struggle or as a disease (e.g., *poverty infects the city*). Spanish also uses a high concentration of poverty as a disease, but it also employs war-specific language about poverty, more often than English (e.g., *guerra contra la pobreza* 'war against poverty'). This is interesting given that the *war against X* metaphoric construction was apparently introduced and popularized in American news media and political rhetoric (Elwood, 1995; see also Flusberg, Matlock, & Thibodeau, 2018).

English and Spanish diverge most notably on metaphors that express poverty as physical harm (or dealing with poverty as violent confrontation) or as a state change of an object (growing or shrinking in size). These differences are driven by a few subgroupings. For instance, in Spanish the metaphor POVERTY IS A CHANGING (GROWING) ENTITY is deeply entrenched (more so than in English), and thus, expressions referring to poverty in terms of physical expansion or extension are common (e.g., *aumentar la pobreza extrema* 'augment extreme poverty', *la pobreza se ha extendido* 'poverty has extended'). In English, POVERTY IS A DISEASE is prominent (e.g., *poverty infected the city*) and thus is linguistically realized in a more robust distribution compared to Spanish. In both languages, Location/Motion metaphors are slightly more common than Violence/Harm metaphors, with more high-density subcategories (categories of over 40 instances per 1,000 results). It is useful to examine a few subframes within this category, especially those pertaining to vertical movement. While the two languages show similar concentrations of static location metaphors (*live/dwell in poverty*, *vivir en la pobreza*), as well as metaphors about confinement or containment in a bounded region (*mired in poverty*, *canasta de pobreza* 'container of poverty'), English seems to favor vertical motion metaphors, as in *catapult/leap out of poverty* and *slide/fall into poverty*. As Dodge (2016) observes, there is more emphasis on manner of motion in English, perhaps resulting from the manner-encoding nature of English verbs. Since GOOD IS UP and BAD IS DOWN are common primary metaphors for negative social and psychological states, and since poverty is seen as a bad social or personal circumstance, we expect many metaphors that emphasize being (unwillingly) knocked down into a state of poverty, and the desire to rise from that low state.

483

The high counts of these metaphors for both languages exhibit slight differences that are informative about how emphatically news in each of these languages conveys these metaphors. This invites some questions about the cognitive status of poverty metaphors in the minds of speakers of each language. For instance, given that BAD IS DOWN is less frequent in Spanish, with expressions like *plunge into poverty*, does this mean that Spanish speakers feel the negative connotations of poverty to a lesser extent? These computationally driven corpus findings may shed light on possible experimental directions to investigate how the language being used may influence conceptualization and decision-making.

6.2. CANCER METAPHORS IN English ACROSS TWO CORPORA

Using the same system,[7] and inheriting the same primary metaphor networks as those used in poverty metaphor identification, we see similar patterns emerging for cancer metaphors. The corpora are much smaller and yield comparably fewer results. Also note that the Gigaword corpora used for the Spanish–English poverty comparison were from newswire sources, which are likely to include proportionally more discussion of socioeconomic issues, including poverty. Cancer, on the other hand, is a niche discussion topic. It appears in the news, but more often, texts with a high concentration of metaphoric language about cancer tend to come from blogs and online forums. Using the same source domain semantic categories as for poverty, we supply a within-genre cross-corpus comparison between a general (and large) corpus (filtered for blog data only) and a specialized corpus with only blog and forum data. Table 4 summarizes results surrounding the topic of cancer in English.

The results in Table 4 are presented in order of actual, rather than relative, difference in normalized frequencies between the GLoWbE corpus and the specialized corpus. The presence of mostly negative values for the specialized corpus in the 'Dif.' column indicates that the specialized corpus (predictably) contains a higher quantity of most metaphors for cancer. From this, we infer that more metaphoric language about cancer is employed by writers who are themselves stakeholders in the cancer world, compared to what we would expect in blog texts at large.

[7] They are the same in terms of their source domain frames, but are adjusted to account for a different target domain (Cancer). These are: CANCER IS A JOURNEY, DEALING WITH CANCER IS PHYSICAL COMBAT, CANCER PATIENT IS COMBATANT, CANCER TREATMENT IS GAMBLING, CANCER TREATMENT IS WAR. As discussed before for Poverty, since these specific metaphors connect in their network to higher-level primary metaphors and entailments, linguistic metaphors will be detected even though they do not have explicit metaphors dedicated to them (as described in Figure 1). Therefore, the results in Table 4 reflect a much broader set of metaphors beyond those seeding the system.

484

TABLE 4. *Results for Cancer metaphors in English from a general and a specialized corpus (per 10 results) (quartile ranges: 0.5–4, 0.2–0.5, 0.1–0.2, 0–0.1). NF: normalized frequency; LU: Lexical Unit; Dif.: actual difference*

	GLoWbE NF	Specialized NF	Dif.	Example source LUs
Violence/Harm				
Physical struggle	0.46	2.15	−1.69	*attack, beat, fight, struggle*
Controlling entity	0.36	1.21	−0.85	*hold, grab, grip, tighten*
Burden	0.26	1.03	−0.77	*load, burden, ton, weight*
Physical harm	0.82	1.31	−0.49	*hit, abuse, threat, impact*
Physical competition	0.20	0.47	−0.27	*win, lose, defeat, challenge*
Darkness	0.01	0.19	−0.18	*cloud, shadow*
Destruction	0.37	0.47	−0.10	*force, blow, erosion, crush*
Theft	0.03	0.09	−0.06	*steal, rob, take*
Fire	0.06	0.09	−0.03	*burn, spark*
Crime	0.18	0.19	−0.01	*victim, sentence, condemn*
War	0.40	0.37	0.03	*war, battle, conquer, weapon*
Disease	0.11	–	–	*(cancer as a) plague*
Tyranny	0.01	–	–	*bully*
Category total:	**3.27**	**7.57**		
Location/Motion				
Be located	0.89	3.36	−2.47	*live in/with, borderline*
Forward movement	1.13	1.78	−0.65	*reach, pass, advance, journey*
Path	0.21	0.65	−0.44	*way, avenue, route*
Upward movement	0.38	0.56	−0.18	*climb, lift, boost, raise*
Enablements to motion	0.12	0.28	−0.16	*(open) door, portal, pave (way)*
Gap	0.09	0.19	−0.10	*gap, divide*
Downward movement	0.54	0.47	0.07	*drop, fall, decline, slip*
Impediments to motion	0.48	0.37	0.11	*impede, suppress, hinder*
Confinement	0.41	0.09	0.33	*isolate, trap, imprison, wall*
Level	0.12	–	–	*level*
Backward movement	0.01	–	–	*fall back, retreat*
Category total:	**4.38**	**7.75**		
Object/entity properties				
Growth in size	0.16	1.21	−1.05	*grow, form, swell, inflate*
Game	0.09	0.56	−0.47	*roulette, bet, (cancer) card*
Plants	0.13	0.56	−0.43	*stem, root, seed, plant*
Reduction in size	0.47	0.56	−0.09	*cut, shrink, slash*
Monster	0.04	0.09	−0.05	*evil, fiend, villain, beast*
Machine operations	0.17	0.19	−0.02	*repair, malfunction, operate*
Features of a building	0.65	0.47	0.18	*build, collapse, stable, pillar*
Eating	0.27	0.19	0.08	*consume, eat (away at)*
Animal	0.09	–	–	*ravage, vicious, predator*
Pest	0.01	–	–	*infest*
Entity observed	0.01	–	–	*be blind to, notice*
Liquid	0.04	–	–	*flood, cascade, wave, flow*
Category total:	**1.45**	**2.43**		

485

The quartile color concentrations in Table 4 illustrates how genres dedicated to cancer talk, such as blogs and fora, show greater frequency of the two metaphor families, Violence/Harm and Location/Motion, which dominate discussions of cancer more generally. Motion metaphors concentrated around discussions of moving forward (*cancer journey*, *the path forward*), dealing with obstacles on a path (*get through this*, *overcome cancer*), and being collocated with cancer (*live with cancer*). Among Violence and Harm metaphors, cancer is either an enemy or antagonistic entity that attacks, beats, or hits the patient, or it is a more abstract force-dynamic antagonist that grips, holds, or squeezes the patient. In the first case, the patient is an equal opponent who can fight back (and potentially win the war), while in the second case, the patient is demoted to a helpless entity that is manipulated and held down. Aside from the two dominant metaphor groups, there are also some spikes in the source domain of object states, especially changes in object size. This is due to the frequent description of cancers as growing, spreading (or conversely, shrinking or diminishing) entities inside the body or in society at large.

This latter finding points to an important observation we make about conceptual categorization in cancer-related metaphor. Namely, there is frequent compartmentalization into three different levels in the target domain: the societal level, in which cancer treatment policies and funding are of primary importance; the individual level, in which cancer is reified as a person that the patient 'fights' or 'travels with'; and the physiological level, at which cancer cells are seen as moving, invading, or attacking within the body (see Semino et al., 2004, for discussion of the last type). Sentences (9), (10), and (11) illustrate these three perspectives on the metaphoric description of cancer.

(9)   The ICR's mission is to make the discoveries that *defeat* cancer. (societal) (CAN_SPEC_1052)

(10) My own brand of faith simply provided me the will to *endure in the face* of cancer's randomness. (personal) (CAN_SPEC_4)

(11) Kris Carr has *halted* tumors growing in her liver for over 7 years by focusing on nutrition. (physiological) (GLO_3914)

This difference is important because it highlights the presence of slightly different metaphor systems from what we observed for the domain of poverty, or other social problems. Both poverty and cancer are personal problems that are metaphorically construed as opponents (enemies) or states to travel through and out of, the two most frequently used metaphors in both domains. But cancer also has a physical, physiological component that is relevant to the level of the individual (Semino et al., 2004, p. 1279). We add that, although cancer is a physical property of the body and 'grows' within the body, it usually cannot be readily perceived via the senses, and therefore the physical bodily effects of cancer are more-or-less metaphorically construed.

486

7. Discussion

The metaphor identification system made a large variety of data instantly available in two languages. By analyzing specific sentences in the output, we can see that although the movement-related words (*slide*, *fall*, *rise*) and violence words (*battle*, *attack*, *fight*) may both be used in metaphoric expressions about both poverty and cancer, the metaphors motivating their use differ in the two domains. Cancer is talked about as a journey, with paths and impediments (rarely the representation of poverty); the motivation in the cancer journey is to arrive at a healthy state (location), while movement in poverty is motivated by trying to flee a bad (often low) location. For the cancer patient, this is an extension of the already entrenched metaphor LIFE IS A JOURNEY, where the experience of cancer is a new phase in the already-ongoing journey of life. This integration of the cancer experience with the life journey is already attested in English narratives by sufferers of breast cancer (Gibbs & Franks, 2002). The journey frame is evoked with not only the lexical item *journey*, but also constructions involving paths (*road to cancer recovery*), obstacles (*a bump in the road*), and enablements (*cancer treatment is going smoothly*) (see also Magaña & Matlock, 2018). Construals of poverty (and socioeconomic states in general) as specifically a journey are fairly uncommon. This is perhaps because it is less common to associate one's personal socioeconomic status with LIFE IS A JOURNEY than with one's own health state, as is the case for CANCER IS A JOURNEY. Nevertheless, some instances of poverty as a journey do appear, as in (12).

(12) Her *journey out of* poverty has been marked by single-mindedness and luck. (ENGW_114719)

Although (12) is a journey metaphor, it conveys the vertical motion typical of poverty metaphors, via the construction *NOUN out of X*. This reflects how discussions of poverty focus less on forward motion and more on vertical motion, where poverty is often a low location that one inadvertently gets into (or is forced into) and wants to escape. This is true in both languages, but more so in English. The low location is also sometimes confining, drawing together not only ACTION IS MOTION, but also BAD IS DOWN and STATES ARE LOCATIONS. This fusion of multiple primary metaphors is exemplified in (13) and (14).

(13) The Congo we're talking about has *mountains* of debt and *canyons* of poverty, so it's a candidate for the debt cancellation that was trumpeted last year. (ENGW_113705)

(14) But it does not necessarily lead us to a more upwardly mobile middle class or rescue those *drowning* in poverty. (ENGW_109128)

Mountains and canyons are topographic features of the terrain, with extreme height and depth. The mountains of debt, although high, do not

487

employ GOOD IS UP, rather DIFFICULTIES ARE IMPEDIMENTS TO MOTION (mountains are hard to traverse on a journey) as well as MORE IS UP. Canyons evokes BAD IS DOWN and UNCHANGEABLE STATES ARE CONFINING LOCATIONS.

Although cancer is also a negative state, and presumably, should use the BAD IS DOWN entailments of movement metaphors, we never see expressions like *fall into cancer or *live in cancer. The cancer experience is most commonly mapped to the path of movement (*my cancer path*), a co-mover or companion (*live with cancer*), or an obstacle in one's normal life path (*overcome cancer*), and less as a location where one ends up in or tries to move away from. Perhaps this is due to seeing cancer as a more transient state, as health states tend to be, and one that happens to be particularly unwelcome as one proceeds with the normal journey of life.

An enormous caveat here is that many trigger words evoke violence and movement scenarios simultaneously. As the numbers are reported in the current work, it is not immediately clear that some datapoints can be cross-categorized into two or more frame families (indeed, in the calculations we only report each datapoint along with the first metaphor suggested by the metaphor identifier). But lexical items often evoke multiple concrete frames in a complex way, as is the case with the use of *shackles* in (15), and *invasive* in (16), and Spanish *desierto* 'desert' in (17).

(15) Unfair trade rules do not only prevent poor people from throwing off the *shackles* of poverty, but *shackle* poor people and poor communities still further. (ENGW_81227)

(16) The authors of the report provide a picture of the number of cancer survivors who had previously been diagnosed with an *invasive* cancer. (GLO _15306)

(17) Y hay muchas formas de desierto: el *desierto* de la pobreza, el *desierto* del hambre y de la sed. (ESGW_501545722)
'There are many types of deserts: the *desert* of poverty, the *desert* of hunger and of thirst.'

In the system, *shackle* is classified as belonging to a family of frames that express confinement. However, a deeper analysis of the semantics of shackles in (15) would reveal how it should also be categorized as a form of Burden (shackles weigh you down), a type of Harm (shackles injure the body), and a type of Impediment to motion (shackles restrict movement). Similarly, *invasive* in (16) is categorized as Movement, but also used to describe movement of unwanted entities that can create harm, and *invade* is often associated with military encroachment. Finally, a desert (17) is a vast open area through which one can move, and a terrain that can be dangerous or fatal (the first being an ACTION IS MOTION metaphor, the second being a STATES ARE

488

LOCATIONS metaphor). Although the salient semantics of this word is spatial, the metaphorically relevant semantics concern the dangers (and helplessness) of this type of terrain, which might be more relevant in poverty metaphors. The problem of primary and secondary classification is one of polysemy in the metaphoric usage of individual lexical items, discussed in the literature on the challenges of grouping metaphors (Cameron, 2010; Deignan, 2010). This should be considered as this (or other computational metaphor identification) systems are further refined.

In this study, we illustrated the effectiveness of an automated method for finding a large number and broad variety of metaphoric expressions across two domains and two languages. These discoveries were made possible from the inclusion of high-level primary metaphors in the metaphor database that propel the automated identification system. Cancer and poverty metaphors were found by means of primary metaphors in the higher levels of the metaphor networks, of which more specific metaphors such as POVERTY IS A DISEASE and CANCER IS PHYSICAL COMBAT, among others, are subcases.

MetaNet allows researchers to identify metaphor in very large texts, and across languages. It may not provide as broad a coverage as some automated metaphor identification systems, but it can be used to query particular target domains, or domains of knowledge in which we would like to observe distributions of metaphoric language, and to do so cross-linguistically. Though not exhaustive, the results provide insights that could lead to further quantitative and qualitative analyses. Its large data output yields a good starting point for understanding conceptual similarities and differences in the domain(s) of interest. Other domains can be added to the core primary metaphor network with minimal tweaks, mostly in the form of adding subcase metaphors and additional lexical items and frames. This iterative process relies on a computationally operationalized version of primary metaphor networks and semantic frames as the source and target domains of conceptual metaphors. We have thus presented one example of MetaNet's pipeline and a selection of results from the domains of cancer and poverty as proof of concept, hoping that such a system will make a valuable addition to metaphor analysts' computational linguistic toolkit.

REFERENCES

Barsalou, L. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition* **10**(1), 82–93.

Birke, J. & Sarkar, A. (2006). A clustering approach for the nearly unsupervised recognition of nonliteral language. Paper presented at the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006).

Burnes, S. (2011). Metaphors in press reports of elections: Obama walked on water, but Musharraf was beaten by a knockout. *Journal of Pragmatics* **43**(8), 2160–2175.

Cahn, J. & Cahn, E. (1964). The war on poverty: a civilian perspective. *Yale Law Journal* **73**(8), 1317–1352.

Cameron, L. (2010). 'What is metaphor and why does it matter?' in L. Cameron & R. Maslen (Eds.), *Metaphor analysis: research practice in applied linguistics, social sciences and the humanities* (pp. 3–25). London: Equinox.

Casarett, D., Pickard, A., Fishman, J. M., Alexander, S. C., Arnold, R. M., Pollak, K. I. & Tulsky, J. A. (2010). Can metaphors and analogies improve communication with seriously ill patients? *Journal of Palliative Medicine* **13**(3), 255–260.

David, O. (2017). Computational approaches to metaphor: the case of MetaNet. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 574–589). Cambridge: Cambridge University Press.

David, O. A. (2016). Metaphor in the grammar of argument realization. Unpublished doctoral dissertation, University of California, Berkeley.

David, O. A., Lakoff, G. & Stickles, E. (2016). Cascades in metaphor and grammar: a case study of metaphors in the gun debate. *Constructions and Frames* **8**(2), 165–213.

Davies, M. (2013). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries*. Available online at <http://corpus.byu.edu/glowbe/>.

Davies, M. & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* **36**(1), 1–28.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam/Philadelphia: John Benjamins.

Deignan, A. (2010) The cognitive view of metaphor: Conceptual Metaphor Theory, in L. Cameron & R. Maslen (Eds.), *Metaphor analysis: research practice in applied linguistics, social sciences and the humanities* (pp. 44–56). London: Equinox.

Demjén, Z., Semino, E. & Koller, V. (2016). Metaphors for 'good' and 'bad' deaths. *Metaphor and the Social World* **6**(1), 1–19.

Demmen, J., Semino, E., Demjén, Z., Koller, V., Hardie, A., Rayson, P. & Payne, S. (2015). A computer-assisted study of the use of Violence metaphors for cancer and end of life by patients, family carers and health professionals. *International Journal of Corpus Linguistics* **20**(2), 205–231.

Do Dinh, E.-L. & Gurevych, I. (2016). Token-level metaphor detection using neural networks. *Proceedings of the Fourth Workshop on Metaphor in NLP* (June) (pp. 28–33). Online: <http://www.aclweb.org/anthology/W16-1104>.

Dodge, E. K. (2016). A deep semantic corpus-based approach to metaphor analysis. *Constructions and Frames* **8**(2), 256–294.

Dodge, E. K., Hong, J. & Stickles, E. (2015). MetaNet: deep semantic automatic metaphor analysis. *Proceedings of the Third Workshop on Metaphor in NLP* (pp. 40–49). Denver, Colorado, 5 June 2015. Association for Computational Linguistics. Online: <http://www.aclweb.org/anthology/W15-1405>.

Dunn, J. (2013a). Evaluating the premises and results of four metaphor identification systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7816 LNCS(PART 1)* (pp. 471–486). Online: <https://link.springer.com/chapter/10.1007/978-3-642-37247-6_38>.

Dunn, J. (2013b). What metaphor identification systems can tell us about metaphor-in-language. *Proceedings of the First Workshop on Metaphor in NLP, Atlanta Georgia, 13 June 2010* (pp. 1–10). Online: <http://www.aclweb.org/anthology/W13-0901>.

Elwood, W. N. (1995). Declaring war on the home front: metaphor, presidents and the war on drugs. *Metaphor and Symbol* **10**(2), 93–114.

Fellbaum, C. (1998). *WordNet: an electronic lexical database*. Cambridge: MIT Press.

Fillmore, C. J. & Atkins, B. T. (1992). Toward a frame-based lexicon: the semantics of RISK and its neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: new essays in semantic and lexical organization* (pp. 75–102). New York/London: Routledge.

Fillmore, C. J., Johnson, C. R. & Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography* **16**(3), 235–250.

Flusberg, S. J., Matlock, T. & Thibodeau, P. H. (2017). Metaphors for the war (or race) against climate change. *Environmental Communication* **11**(6), 769–783.

Flusberg, S. J., Matlock, T. & Thibodeau, P. H. (2018). War metaphors in public discourse. *Metaphor and Symbol* **33**, 1–18.

490

Gedigian, M., Bryant, J., Narayanan, S. & Ciric, B. (2006). Catching metaphors. *Proceedings of the Third Workshop on Scalable Natural Language Understanding ScaNaLU 06* (June), (pp. 41–48). Online: <http://www1.icsi.berkeley.edu/~jbryant/GedigianBryantNarayananMetaphor.pdf>.

Gibbs, R. W. (2015). Counting metaphors: What does this reveal about language and thought? *Cognitive Semantics* **1**, 155–177.

Gibbs, R. W., Jr. (1994). *The poetics of mind: figurative thought, language, and understanding*. Cambridge: Cambridge University Press.

Gibbs, R. W., Jr. & Franks, H. (2002). Embodied metaphor in women's narratives about their experiences with cancer. *Health Communication* **14**(2), 139–165.

Gibbs, R. W., Jr., Lima, P. L. C. & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics* **36**(7), 1189–1210.

Gordon, J., Hobbs, J. R., May, J., Morbini, F. & Vista, P. (2015). High-precision abductive mapping of multilingual metaphors. *Proceedings of the Third Workshop on Metaphor in NLP* (**2**), 50–55. Online: <http://www.aclweb.org/anthology/W15-1406>.

Grady, J. E. (1997). Foundations of meaning: primary metaphors and primary scenes. Unpublished doctoral dissertation, University of California, Berkeley.

Gutiérrez, D. E., Shutova, E., Marghetis, T. & Bergen, B. K. (2016). Literal and metaphorical senses in compositional distributional semantic models. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, August 7–12, 2016* (pp. 183–193). Online: <http://www.aclweb.org/anthology/P16-1018>.

Hong, J. (2016). Automatic metaphor detection using constructions and frames. *Constructions and Frames* **8**(2), 293–320.

Johnson, M. (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason*. Chicago/London: University of Chicago Press.

Kövecses, Z. (2005). *Metaphor in culture: universality and variation*. Cambridge: Cambridge University Press.

Kövecses, Z. (2015). *Where metaphors come from: reconsidering context in metaphor*. Oxford: Oxford University Press.

Krishnakumaran, S. & Zhu, X. (2007). Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language* (pp. 13–20). Association for Computational Linguistics. Online: <https://dl.acm.org/citation.cfm?id=1611531>.

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about thought*. Chicago: University of Chicago Press.

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in Cognitive Science* **4**, 773–785.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Lakoff, G. & Kövecses, Z. (1987). The cognitive model of anger inherent in American English. In N. Quinn (Ed.), *Cultural models in language and thought* (pp. 195–221). Cambridge: Cambridge University Press.

Lederer, J. (2013). Assessing claims of metaphorical salience through corpus data. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 1255–1260). Austin, TX: Cognitive Science Society.

Levin, L., Mitamura, T., Fromm, D., MacWhinney, B., Carbonell, J., Feely, W., Frederking, R., Gershman, A. & Ramirez, C. (2014). Resources for the detection of conventionalized metaphors in four languages. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (pp. 498–501). Online: <https://pdfs.semanticscholar.org/1b08/69f556fc45c935b239447929b121762cac98.pdf>.

Lönneker, B. (2003). Is there a way to represent metaphors in WordNets? Insights from the Hamburg Metaphor Database. *Proceedings of the ACL 2003 Workshop on Lexicon and Figurative Language – Volume 14* (pp. 18–27). Online: <https://dl.acm.org/citation.cfm?id=1118978>.

MacWhinney, B. & Fromm, D. (2014). Two approaches to metaphor detection. *Proceedings of the 9th Edition of the Language, Resources and Evaluation Conference (LREC 2014)* (pp. 2501–2506). Online: <https://pdfs.semanticscholar.org/e8bc/8b3eeca8fe7146c1d830c4fff495c05b6568.pdf>.

Magaña, D. & Matlock, T. (2018). How Spanish speakers use metaphor to describe their experiences with cancer. *Discourse & Communication*. Available online <https://doi.org/10.1177/1750481318771446>.

Martin, J. H. (1988). A computational theory of metaphor. Unpublished doctoral dissertation, University of California, Berkeley.

Martin, J. H. (1994). MetaBank: a knowledge-base of metaphoric language conventions. *Computational Intelligence* **10**(2), 134–149.

Martin, J. H. (2006). A corpus-based analysis of context effects on metaphor comprehension. In S. T. Gries & A. Stefanowitsch (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 214–236). Berlin: Mouton de Gruyter.

Mason, Z. J. (2004). CorMet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics* **30**(1), 23–44.

Mendonça, A., Jaquette, D., Graff, D. & DiPersio, D. (2011). *Spanish Gigaword second edition (LDC2011T12)*. Philadelphia: Linguistic Data Consortium.

Mohler, M., Tomlinson, M. & Rink, B. (2015). Cross-lingual semantic generalization for the detection of metaphor. *International Journal of Computational Linguistics and Applications* **6**(2), 117–140.

National Cancer Institute (2017). *National Cancer Act of 1971*. Retrieved from <https://dtp.cancer.gov/timeline/flash/milestones/M4_Nixon.htm>.

Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N. & Frieder, O. (2013). Metaphor identification in large texts corpora. *PLoS ONE* **8**(4), 1–9. Available online <https://doi.org/10.1371/journal.pone.0062343>.

Olweny, C. L. M. (1997). Effective communication with cancer patients: the use of analogies – a suggested approach. *Annals of the New York Academy of Sciences* **809**, 179–187.

Parker, R., Graff, D., Kong, J., Chen, K. & Maeda, K. (2011). *English Gigaword fifth edition (LDC2011T07)*. Philadelphia: Linguistic Data Consortium.

Philip, G. (2004). Locating metaphor candidates in specialized corpora using raw frequency and keyword lists. In F. MacArthur, J. L. Oncins-Martínez, M. Sánchez-García & A. M. Piquer-Píriz (Eds.), *Metaphor in use: context, culture, and communication* (pp. 85–105). Amsterdam: John Benjamins.

Pragglejaz Group (2007). MIP: a method for identifying metaphorically used words in discourse. *Metaphor and Symbol* **22**(1), 1–39.

Radden, G. (2011). Spatial time in the West and the East. In M. Brdar, M. Omazić, V. P. Takač, T. Gradečak-Erdeljić & G. Buljan (Eds.), *Space and time in language* (pp. 1–400). Frankfurt: Peter Lang.

Ruppenhofer, J. K., Ellsworth, M., Petruck, M. R. L., Baker, C. F. & Scheffczyk, J. (2016). *FrameNet II: extended theory and practice*. Berkeley, CA: International Computer Science Institute.

Semino, E. (2010). Descriptions of pain, metaphor, and embodied simulation. *Metaphor and Symbol* **25**, 205–226.

Semino, E., Demjén, Z., Demmen, J., Koller, V., Payne, S., Hardie, A. & Rayson, P. (2015). The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ Supportive & Palliative Care* **7**(1), 1–7.

Semino, E., Heywood, J. & Short, M. (2004). Methodological problems in the analysis of metaphors in a corpus of conversations about cancer. *Journal of Pragmatics* **36**(7), 1271–1294.

Semino, E. & Masci, M. (1996). Politics is football: metaphor in the discourse of Silvio Berlusconi in Italy. *Discourse & Society* **7**(2), 243–269.

Shutova, E. (2010). Models of Metaphor in NLP. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (July)* (pp. 688–697). Online: <https://dl.acm.org/citation.cfm?id=1858752>.

Shutova, E. & Sun, L. (2013). Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Proceedings of NAACL-HLT 2013, Atlanta, Georgia, 9–14 June 2013* (pp. 978–988). Online: <http://www.aclweb.org/anthology/N13-1118>.

Shutova, E., Teufel, S. & Korhonen, A. (2012). Statistical metaphor processing. *Computational Linguistics* **39**(2), 301–353.

Small-McKinney, A. (2014). I Am Strong, I Am Not (The Breast Cancer Journey). Retrieved from <http://community.breastcancer.org/blog/i-am-strong-i-am-not/>.

Steen, G. J. (1999). From linguistic to conceptual metaphor in five steps. In R. W. Gibbs & G. J. Steen (Eds.), *Metaphor in cognitive linguistics* (pp. 57–77). Amsterdam/Philadelphia: John Benjamins.

Steen, G. J., Biernacka, E., Dorst, A. G., Kaal, A. A., López-Rodríguez, I. & Pasma, T. (2010a). Pragglejaz in practice: finding metaphorically used words in natural discourse. In G. Low, Z. Todd, A. Deignan & L. Cameron (Eds.), *Researching and applying metaphor in the real world* (pp. 165–184). Amsterdam/Philadelphia: John Benjamins.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A. & Krennmayr, T. (2010b). Metaphor in usage. *Cognitive Linguistics* **21**, 765–796.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T. & Pasma, T. (2010c). *A method for linguistic metaphor identification: from MIP to MIPVU*. Amsterdam: John Benjamins.

Stefanowitsch, A. (2006). Words and their metaphors: a corpus-based approach. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 63–105). Berlin/NewYork: Mouton de Gruyter.

Stefanowitsch, A. & Gries, S. Th. (Eds.) (2006). *Corpus based approaches to metaphor and metonymy*. Berlin/New York: Mouton de Gruyter.

Stewart, M. (2014). The road to pain reconceptualisation: Do metaphors help or hinder the Journey? *Pain and Rehabilitation: The Journal of Physiotherapy Pain Association* **36**, 24–31.

Stickles, E., David, O., Dodge, E. K. & Hong, J. (2016). Formalizing contemporary conceptual metaphor theory. *Constructions and Frames* **8**(2), 166–213.

Sullivan, K. (2016). Integrating constructional semantics and conceptual metaphor. *Constructions and Frames* **8**(2), 141–165.

Sullivan, K. S. (2006). Frame-based constraints on lexical choice in metaphor. In *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society* (*Vol. 32*) (pp. 387–400). Online: <https://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/3476/3177>.

Sweetser, E., David, O. & Stickles, E. (in press). MetaNet: automated metaphor identification across languages and domains. In Bolognesi, M., Brdar, M. & Despot, K. S. (Eds.), *Fantastic metaphors and where to find them: traditional and new methods in figurative language research*. Amsterdam: John Benjamins.

Thibodeau, P. H. & Boroditsky, L. (2011). Metaphors we think with: the role of metaphor in reasoning. *PLoS One* **6**(2), 1–11. Retrieved from <https://doi.org/10.1371/journal.pone.0016782>.

Thibodeau, P. H. & Boroditsky, L. (2013). Natural language metaphors covertly influence reasoning. *PLoS ONE* **8**(1), 1–7. Retrieved from <https://doi.org/10.1371/journal.pone.0052961>.

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E. & Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)* (pp. 248–258). Online: <http://www.aclweb.org/anthology/P14-1024>.

Weiss, M. (1997). Signifying the pandemics: metaphors of AIDS, cancer, and heart disease. *Medical Anthropology Quarterly* **11**(4), 456–476.

Wilks, Y. (1975). A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence* **6**, 53–74.

Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence* **11**(3), 197–223.