Building the MetaNet metaphor repository: The natural symbiosis of metaphor analysis and construction grammar

Oana David, oanadavid@berkeley.edu
Ellen Dodge, edodge@icsi.berkeley.edu
Elise Stickles, elstickles@berkeley.edu
Eve Sweetser, sweetser@berkeley.edu
Jisup Hong jhong@icsi.berkeley.edu

University of California, Berkeley & International Computer Science Institute

ICCG 8, Osnabrück, Germany, Sept. 3-6 2014
Overview

- MetaNet project goals
- System design and specifications
- A step-by-step extraction example
- Future applications
MetaNet Project Goals

1. To identify **linguistic metaphors** (LMs, metaphoric phrases) in real texts.

2. Subsequently to identify **conceptual metaphors** (CMs) that those linguistic metaphors evoke.

3. To create a computational system that uses known interconnected metaphor networks to find new LMs, and correctly identify their CMs, based on associated lexical units (LUs) found in the LMs.
Role of constructions

- To identify metaphors in text, grammatical context is needed.
  - Consider ‘poverty cripples the inner city’ vs. ‘we must cripple poverty with new policies’

Resources

- Using MetaNet, FrameNet
  - MetaNet schemas (definitions, relations, roles, and lexical items) designed in large part on the FrameNet model with many commonalities.
  - Large MetaNet growing database of metaphoric and other schema-to-schema relations
  - Perform metaphor extraction also with the help of FrameNet frames and WordNet.

- Additions
  - Metaphors: schema-to-schema relations (frame-to-frame).
  - Constructions: currently simple, but gradually building compatibility with ECG.
  - Schema and construction design is geared towards linguistic metaphor detection in natural texts.
Manual portion

- A database of manually-entered **schemas** (frames and cogs) and **metaphors** (schema-to-schema mappings)
- Input method used is the MetaNet Wiki (4 languages).

<table>
<thead>
<tr>
<th></th>
<th>Persian</th>
<th>Russian</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaphors</td>
<td>190</td>
<td>124</td>
<td>541</td>
<td>732</td>
</tr>
<tr>
<td>schemas</td>
<td>306</td>
<td>265</td>
<td>443</td>
<td>627</td>
</tr>
</tbody>
</table>

- Schemas are interrelated in networks with relations (makes use of, is a subcase of, is a perspective on)
- Schemas are assigned to source and target domain fields of metaphors.
- Schemas have roles as well as lexical units, facilitating detection in texts.
Schema and Metaphor Formalization

Schema1 schema:
- schema_role_1
- schema_role_2
- x-schema

LUs: lexeme1, lexeme2, lexeme3...

TARGET1 IS SOURCE1

Target1 schema:
- target1_role1
- target1_role2
- target1_role3

Source1 schema:
- source1_role1
- source1_role2
- source1_role3

Arrows indicate correspondences between roles in the two schemas.
Metaphor 1

Metaphor 2
Metaphor 1

Metaphor 2
Combining constructions with MetaNet metaphors

- The constructions that the system uses are manually specified simple (two-slot) cxns:
- Their slots are specified for Target or Source status, given commonly-found patterns in that language.
  - Subject(T)-Verb(S): poverty attacks, debt crushes, taxes infect, taxes destroy, wealth cushioned
  - Verb(S)-Object(T): attack poverty, raising taxes, build wealth
  - Noun(T)-noun(S): income gap, poverty slump, debt burden, poverty epidemic, legislative firestorm
  - Noun(S)-of-noun(T): abyss of poverty, hurricane of taxes, dungeon of poverty
  - Adjective(S)-noun(T): crumbling tax system, crippling poverty, oppressive debt
Combining constructions with MetaNet metaphors

• E.g. “crippling poverty”
  • Adjective-Noun (specifically *predicative adjectival* construction, Sullivan 2007)
Metaphor: ECONOMIC HARDSHIP IS PHYSICAL HARM

- **Economic hardship**
  - experiencer
  - state_of_economic_hardship
  - effect_of_economic_hardship

- **Physical harm**
  - harmed_entity
  - cause_of_physical_harm
  - effect_of_harm

Arrow connections indicate:
- poverty
- crippling
System Design

- **Automated portion**
  - A metaphor LM extraction system that retrieves LMs from texts from the wild (corpora, the internet)

- The extraction system uses grammatical constructions to match LMs to CMs.
Step-by-step extraction

• **Step 1:** Extractor finds a target LU in a portion of text.

  *Poverty* continues to cripple millions of Canadians - the majority of which are children - across the country today. ([http://kathydobson.ca/tag/montreal](http://kathydobson.ca/tag/montreal/))

  • Identifies **poverty.n** as an LU in the Poverty schema.

• **Step 2:** using POS tagging recognition and dependency relations, the extractor identifies the cxn that the LU is found in, and its role in that cxn.

  • ‘poverty’ is identified as the subject of a **Subj-Verb** cxn, where subject is TARGET and verb is SOURCE.
Step-by-step extraction

• **Step 3**: Given the identified cxn, the extractor uses POS tagging and dependency relations to identify a source LU in the available constructional slot.
  - poverty continues to cripple (‘cripple’ is the verb in the Subj-Verb cxn, where subject is TARGET and verb is SOURCE).

• **Step 4**: the system matches the identified source LU to a repository schema
  - cripple.v is found in both Harm_to_living_entity and Motion_impediments schemas
Poverty continues to cripple millions of Canadians— the majority of which are children— across the country today.

target LU=**Poverty**
target schema=Poverty
source LU=**cripple**

source schemas=Harm_to_living_entity, Motion_impediments

CMs=POVERTY IS PHYSICAL HARM,
EXPERIENCING A NEGATIVE STATE IS EXPERIENCING HARM,
ECONOMIC HARDSHIP IS PHYSICAL HARM

cxn=T-subj_aspV_S-verb
Poverty continues to *cripple* millions of Canadians— the majority of which are children— across the country today.

**POS:** Poverty=poverty=NN=0 continues=continue=VBZ=1 to=to=TO=2 cripple=cripple=VB=3 millions=million=NNS=4 of=of=IN=5 Canadians=Canadians=NP=6 the=the=DT=7 majority=majority=NN=8 of=of=IN=9 which=which=WDT=10 are=be=VBP=11 children=children=JJ=12 across=across=IN=13 the=the=DT=14 country=country=NN=15 today=today=NN=16 .=.=SENT=17

**Dependency relations:**

```
"Poverty"-0 is ncsbj of "continues"-1
"cripple"-3 is xcomp of "continues"-1
"millions"-4 is dobj of "cripple"-3
"of"-5 is ncm of "are"-11
"Canadians"-6 is dobj of "of"-5
"the"-7 is det of "majority"-8
"majority"-8 is ncsbj of "are"-11
"of"-9 is iobj of "majority"-8
"which"-10 is dobj of "of"-9
"are"-11 is ccomp of "cripple"-3
"children"-12 is xcomp of "are"-11
"across"-13 is ncm of "children"-12
"the"-14 is det of "country"-15
"country"-15 is dobj of "across"-13
"today"-16 is ncm of "are"-11
```
But their crusade against gun control benefits from the hysteria and paranoia that such reckless, inflammatory rhetoric incites.
Step 5: Database Search -- Repository metaphors are used to narrow down the schema selection.

- Is Harm_to_living_entity the source schema for a metaphor where Poverty is the target schema? Is there a POVERTY IS HARM TO LIVING ENTITY metaphor?
  - No.

- Is Motion_impediments the source schema for a metaphor where Poverty is the target schema? Is there a POVERTY IS A MOTION IMPEDIMENT metaphor?
  - No.
Step 5: Database Search --

- Holding the source schema constant, if no metaphor exists where precisely Poverty is the target schema, is there one containing either of the two candidate source schemas where a parent node of Poverty acts as the target?
  - Yes. **ECONOMIC HARDSHIP IS PHYSICAL HARM.** (T:Economic hardship, S: Physical_harm)

- Holding the target constant, if no metaphor exists where either of precisely these two source schemas exist, are there any metaphors where parents of these two source schemas are the source for the Poverty metaphor?
  - Yes. **POVERTY IS PHYSICAL HARM** (T: Poverty, S: Physical_harm)

- Are there any metaphors where both target and source schemas as parents of the schemas associated with the extracted LM?
  - Yes. **EXPERIENCING A NEGATIVE STATE IS EXPERIENCING HARM.** (T:Experiencing_a_negative_state, S: Physical_harm)
Poverty continues to cripple millions of Canadians—the majority of which are children—across the country today.
Summary

- The system relies on existing schema and metaphor networks to produce viable candidate CMs for novel LMs.

- Inheritance relations are leveraged to identify specific LMs as instances of more general metaphors, even when only a few general metaphors have been entered manually.

- The system uses ‘best fit’ on both target and source sides to yield the candidate metaphors for a given LM.
  - Note that although ‘cripple’ additionally is listed as an LU under Motion_impediments schema, no metaphor with this as a source schema is produced.
The Harm Schema Family

- Carefully designed, theoretically correct schema networks and schema-to-schema relations are crucial for this to happen. E.g:

  ‘avalanche of poverty’
  ‘crushed by poverty’
  ‘impacted by poverty’
Future applications: Constructions in the metaphor repository

- Currently, lexical items act as indexes, letting the extractor know what schema to look for (they are associated with a schema).

- Lexical items should be associated with **schema roles** rather than with schemas, for more precise mappings between grammatical slots and mapped schema roles:
Economic hardship

- experiencer
- state_of_economic_hardship
- effect_of_economic_hardship

Physical harm

- harmed_entity
- cause_of_physical_harm
- effect_of_harm

poverty

crippling
‘crippling’ alone does not tell us the role of ‘poverty’ in the crippling event. Poverty is the agent of harm causation, and crippling is the effect of that harm.

-Poverty ↔ Cause of economic hardship ↔ Cause_harm_agent
-Impoverished entity ↔ Affectee of economic hardship ↔ Harmed entity
-Effect of being impoverished ↔ [Effect of harm ↔ Becoming crippled]
Harm_to_living_entity ‘is a subcase of’ Physical_harm with LUs: hurt, injure, wound, cripple, maim, torture, stunt, poison, flog, etc.

‘crippling poverty’ (poverty cripples us)
‘we must cripple poverty’ (we cripple poverty)

Cxn slots must know how to link to LUs that evoke specific schema roles in order to distinguish these.
Future applications: Multiple mapping possibilities

At the moment, system cxns are still simple, and finer distinctions cannot yet be made. Sometimes, T and S are not predictable.

• E.g., Two kinds of Adj-N constructions
  • Domain adjective, e.g. ‘economic boom’ (*a boom that is economic):
    • The adjective is TARGET and the noun is SOURCE
  • Predicative adjective, e.g. ‘crippling poverty’ (a poverty that is crippling people)
    • The adjective is SOURCE and the noun is TARGET

• A system where T and S are not fixed in their assignment to cxn slots is needed; it needs to rely on filler-role relations and type constraints to handle instances like this.
Future applications: Layered constructions

- “Crippling poverty grips / threatens the nation.”
  - *crippling poverty*, POVERTY IS PHYSICAL HARM
  - *poverty threatens*, POVERTY IS AN ADVERSARY

- “We need to eradicate crippling poverty.”
  - *crippling poverty* POVERTY IS PHYSICAL HARM
  - *eradicate poverty* POVERTY/SOCIAL PROBLEMS ARE PLANTS/DISEASES

- “To extinguish crippling poverty…”
  - *crippling poverty*, POVERTY IS PHYSICAL HARM
  - *extinguish poverty* POVERTY IS A DESTRUCTIVE NATURAL FORCE (FIRE)

- Constructions are rarely simple two-slot lexical combinations. Creating a system that takes embedded cxns as input to larger cxns is needed to handle Subj-Verb-Obj sequences.
Conclusions

- A metaphor extraction system that pulls metaphoric language from real-world texts need to have both manual and an automated components, including frames, metaphors and grammatical cxns.

- The manual component need not be exhaustive; ‘best fit’ lattice systems that ‘look up’ (or sideways) in a lattice network can help give results for any novel, even creative LM (e.g., *fan the flames of democracy, social unrest sparked democracy in that country*)
Thank you!

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0022. 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.


List of all Schemas

Contents [hide]

1 Manually Analyzed Schemas
2 Automatically Extracted Schemas

Total number of Schema pages: 627

Manually Analyzed Schemas

Create or edit a Schema:

Total number of Schema pages: 627

A
- Ability to act
- Absorption
- Access
- Access to a location
- Access to an object
- Access to education
- Access to knowledge

E cont.
- Environmental restraints
- Equality
- Erosion
- Essence
- Evaluation
- Event
- Evil
- Evil creature

O cont.
- Object propulsion
- Object transfer
- Obscenity
- Open
- Operating a machine
- Opportunities
- Organization
- Organization of power
### Schema: Erosion

<table>
<thead>
<tr>
<th>Description</th>
<th>The process and experience of effects of erosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closest FrameNet Frame(s)</td>
<td></td>
</tr>
<tr>
<td>Maps to IARPA Source Concept(s)</td>
<td><strong>NATURAL PHYSICAL FORCE</strong> (#1:0.04;erode)</td>
</tr>
<tr>
<td>Other aliases</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td>Includes both causal process and experiencing of erosion</td>
</tr>
<tr>
<td>Family</td>
<td>Harm, Damage</td>
</tr>
<tr>
<td>Tags</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Frame</td>
</tr>
<tr>
<td>Cultural Scope</td>
<td>American English, Mexican Spanish</td>
</tr>
<tr>
<td>Cultural Information</td>
<td>checked cultural scope for English and Spanish - Karie July 8 2014</td>
</tr>
</tbody>
</table>
### Roles:

<table>
<thead>
<tr>
<th>Role Name</th>
<th>Definition/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>eroded_entity</td>
<td></td>
</tr>
<tr>
<td>entity</td>
<td></td>
</tr>
<tr>
<td>erosion_process</td>
<td>progressive, slow, ongoing process</td>
</tr>
<tr>
<td>x-schema</td>
<td></td>
</tr>
<tr>
<td>eroding_effect</td>
<td>depletion of content/mass</td>
</tr>
<tr>
<td>eroding_outcome</td>
<td>loss of structural integrity, functionality</td>
</tr>
<tr>
<td>erosion_cause</td>
<td>entity or process, force</td>
</tr>
</tbody>
</table>
## Related Schemas:

<table>
<thead>
<tr>
<th>Current Schema</th>
<th>Erosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation Type</td>
<td>is subcase of</td>
</tr>
<tr>
<td>Related Schema</td>
<td>Destructive natural process</td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
</tr>
</tbody>
</table>

## Relevant Lexical Units:

- **Lemma** erode.v
- **Lemma** erosion.n
- **Lemma** eat away.v
- **Lemma** corrosion.n
- **Lemma** corrode.v
**Role-based cxn-to-schema links**

- These mappings and bindings naturally link to the schematic meanings already present in grammatical cxns.
  - The transitive cxn already sets up the subject as the causal agent and the object as the affectee.
  - Most gerundive adj-noun cxns have a null instantiated affectee and an overt causal agent in the noun slot.

<table>
<thead>
<tr>
<th>Cause_harm_agent</th>
<th>Cause_harm_affectee</th>
<th>Causal_effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crippling_agent</td>
<td>Crippled_entity</td>
<td>Crippling</td>
</tr>
<tr>
<td>Noun</td>
<td>Ø</td>
<td>Adj.</td>
</tr>
</tbody>
</table>