

A construction-driven, MetaNet-based approach to metaphor extraction and corpus analysis

Elise Stickles, Ellen Dodge, and Jisup Hong

UC Berkeley and International Computer Science Institute

CSDL

November 5, 2014

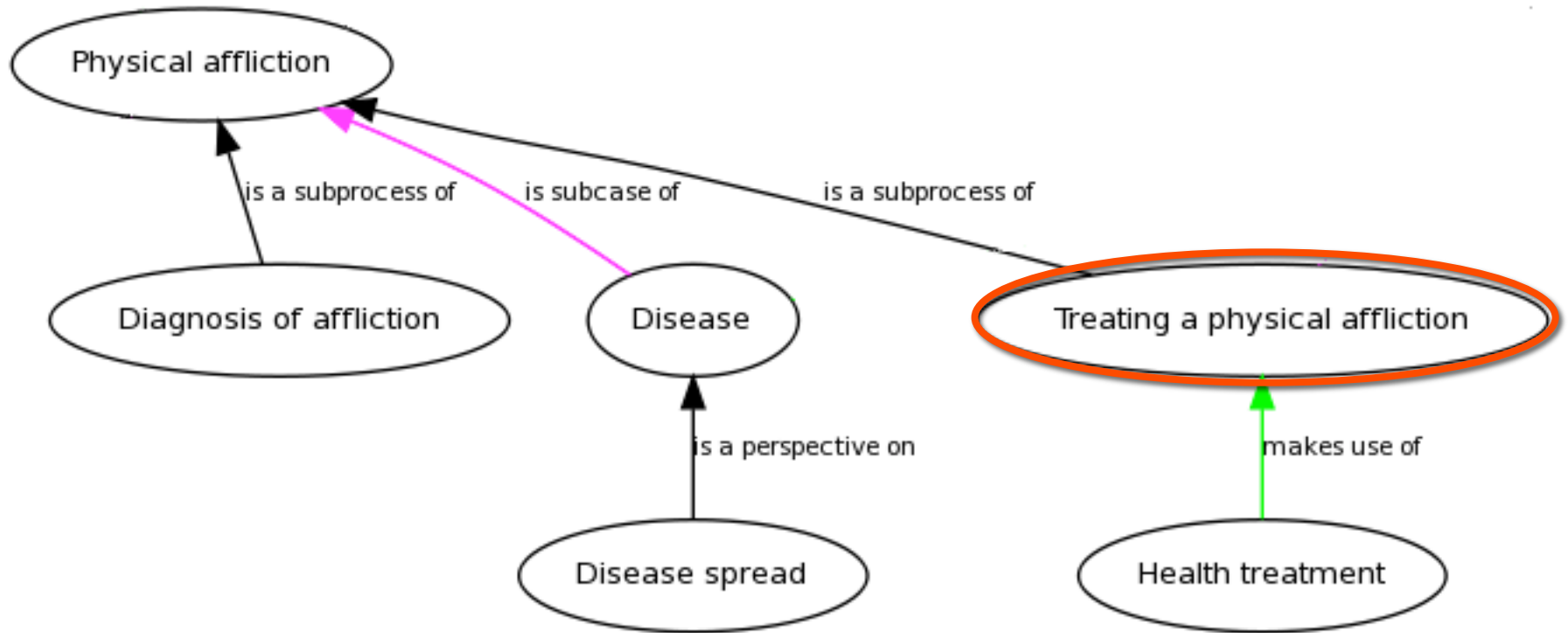
Overview

- MetaNet Analysis project
- Metaphor constructions
- Automated metaphor extraction
- Corpus analysis case study

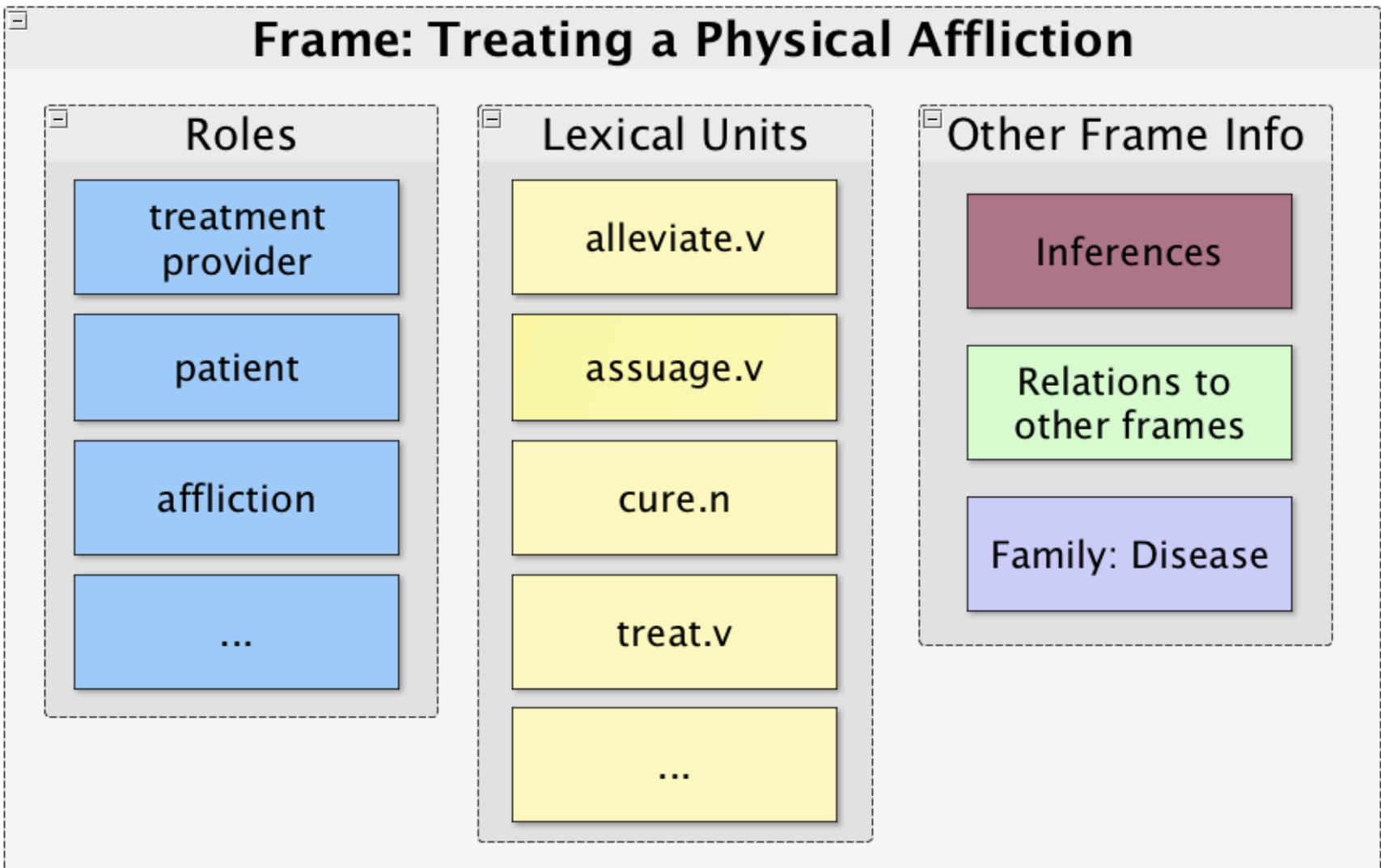
Overview: MetaNet Analysis project

- Goals
 - Development of a multi-lingual repository to support cognitive linguistic analysis of metaphor
 - Automated metaphor extraction system
- Repository
 - Networks of frames and conceptual metaphors
 - Structured relations between frames and metaphors
 - Frames organized into families of semantically coherent domains
 - English repository currently contains 647 frames and 754 metaphors
 - Includes comprehensive coverage of primary metaphors

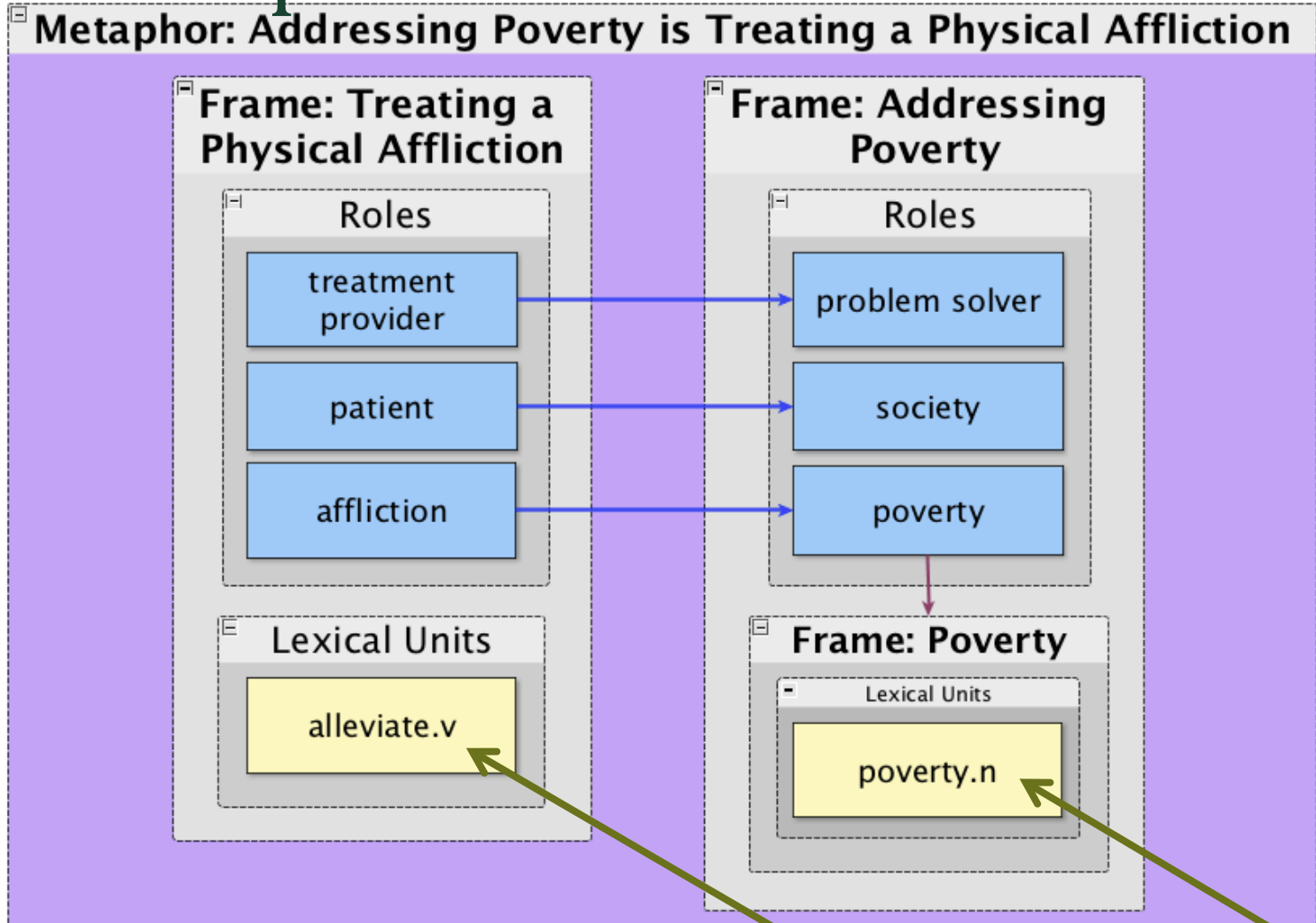
Disease family: network of frames



Frame structure



Metaphor structure



*...neither the free market nor central planning had been able to **alleviate** unemployment and **poverty** (BNC:HKT)*

Metaphoric constructions

- Conceptual metaphors are typically expressed in particular syntactic patterns.
 - Target and source lexemes reliably occupy certain grammatical slots. (Goldberg 1995; Croft 2002; Sullivan 2007, 2013)
 - E.g. Source is verb, Target is argument
- Metaphoric constructions in the MetaNet system are primarily formalized using dependency parse and part of speech tag information

Metaphoric constructions

Metaphoric Construction: Verb-Direct Object

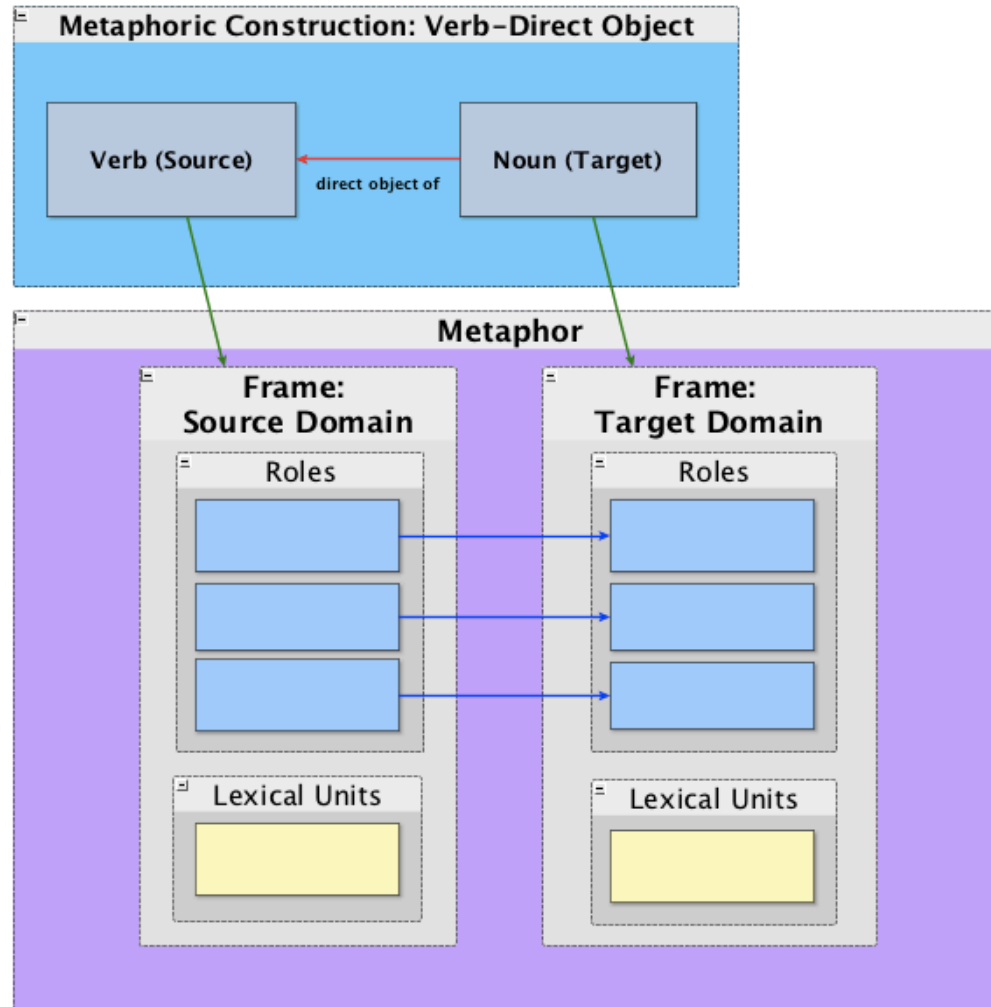
Verb (Source)

direct object of

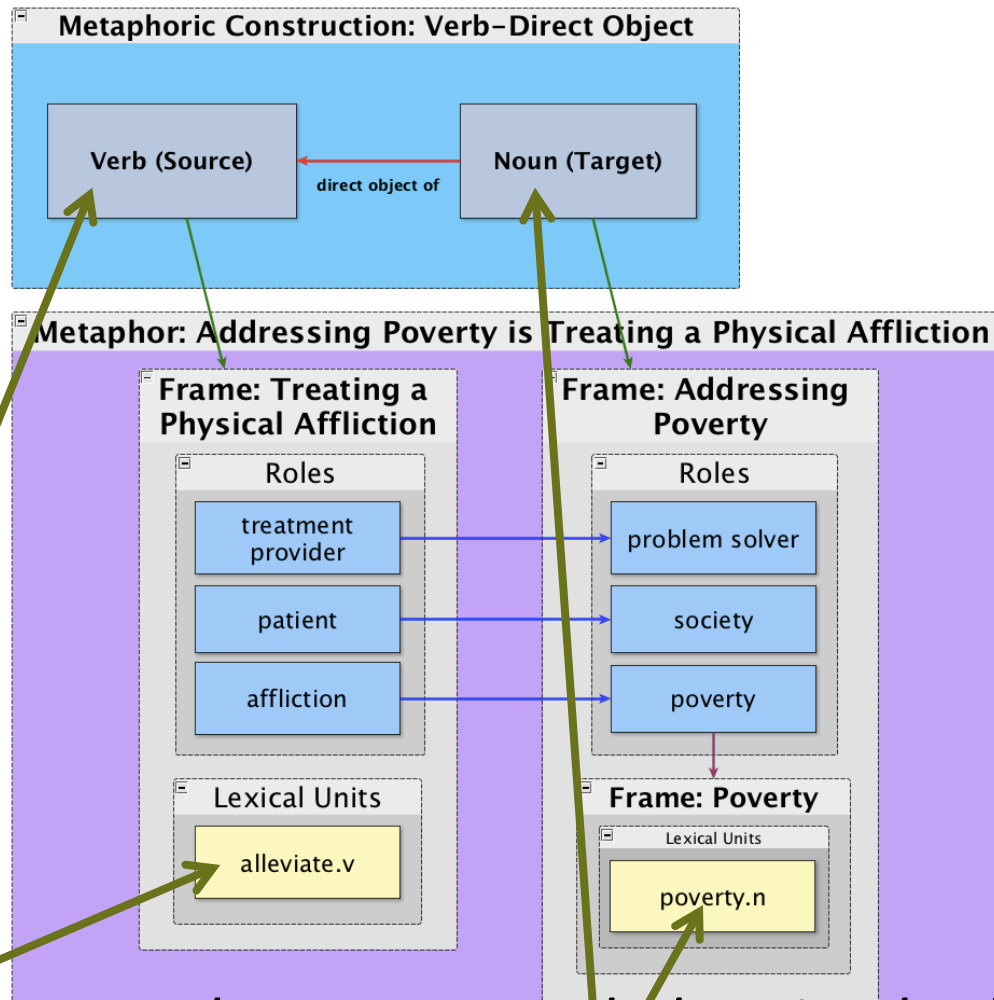
Noun (Target)

*neither the free market nor central planning had been able to **alleviate** unemployment and **poverty** (BNC:HKT)*

Combining metaphor and construction



Combining metaphor and construction



*neither the free market nor central planning had been able to **alleviate** unemployment and **poverty** (BNC:HKT)*

Metaphoric constructions:

Source is verb, Target is argument

- Subject (Target) + Verb (Source)
 - *poverty and early death threatened family life* (BNC:AP7)
- Verb (Source) + Direct Object (Target)
 - *The measure was principally designed to tackle homelessness* (BNC:G20)

Metaphoric constructions:

Source and Target are both nominals

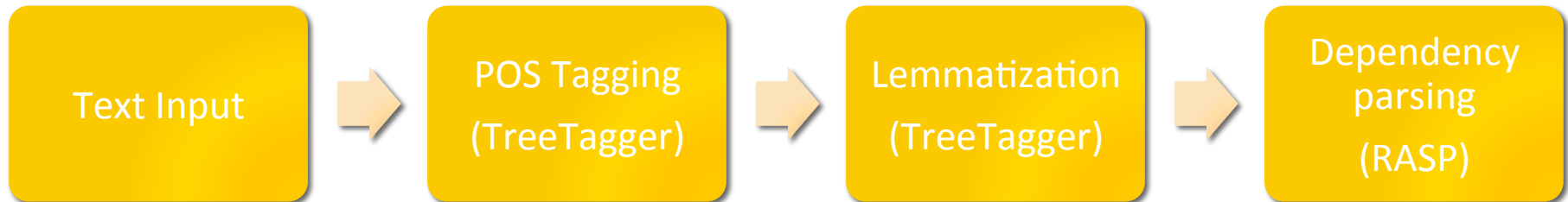
- **N of N: N1 (Source) + of + N2 (Target)**
 - *end the twin **scourges** of mass **poverty** and mass unemployment (BNC:CAK)*
- **N-N compound: N1 (Target) + N2(Source)**
 - *Caught in the **poverty trap**, they are unable to make the savings necessary for business ventures. (BNC:AN3)*

Multi-lingual extraction system

- English, Spanish, Russian, and Persian
 - Taggers:
 - English, Spanish, Russian: TreeTagger
 - Persian: Custom
 - Dependency Parsers:
 - English: RASP
 - Spanish: Freeling
 - Russian/Persian: MALT
- Large corpora
 - Gigaword corpora for English, Spanish, Russian
 - 100 Million+ word corpora for Persian

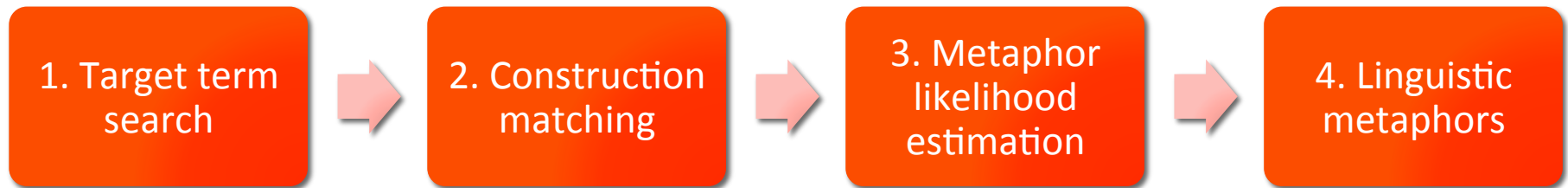
Extraction process overview

- Corpus pre-processing using standard NLP methods



Extraction process overview

- Metaphor extraction using repository and constructions



1. Identify target expressions in the corpus, e.g. *poverty*
2. Construction matching patterns find source expressions grammatically related to target expressions
3. Metaphor candidates (target and source word pairs) are evaluated for metaphoricity
 - Frame relations in the metaphor repository used to determine non-metaphorical relatedness or metaphorical relatedness
4. Filter out low-scoring candidates to create database of linguistic metaphors

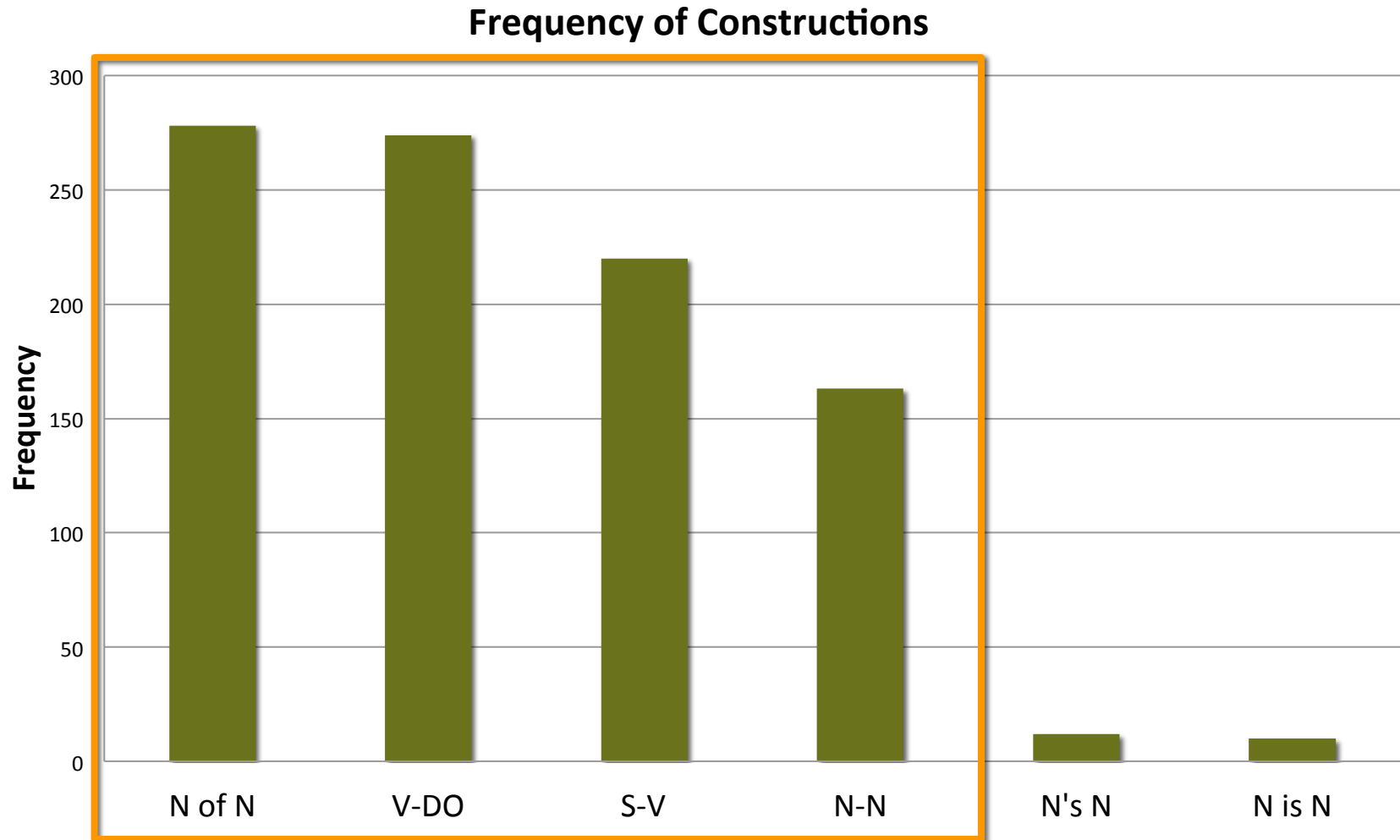
Extraction output

```
{
  "extractor": "CMS",
  "name": "poverty hurt",
  "source": {
    "end": 13,
    "form": "hurts",
    "lemma": "hurt",
    "framenames": [ "Experience pain", "Harm to living entity" ],
    "lpos": "hurt.v",
    "start": 8,
  },
  "score": 0.93,
  "cms": [ "POVERTY_IS_PHYSICAL_HARM",
    "EXPERIENCING_A_NEGATIVE_STATE_IS_EXPERIENCING_HARM" ],
  "cxn": "***T-subj_S-verb",
  "target": {
    "concept": "POVERTY",
    "end": 7,
    "framefamily": "Economic inequality frames",
    "form": "Poverty",
    "lemma": "poverty",
    "start": 0,
    "frame": "Poverty",
  }
}
```


Overview: Case study

- Goal: Illustrate system's potential
 - Large-scale automated extraction
 - Maintain analytic depth
 - Enabled by repository framework and data
- British National Corpus
 - 100 million words
 - Late 20th century (1970s-1990s)
 - 90% written, 10% spoken
- Target domain: Poverty
 - Defined by poverty frame family in repository
 - Extraction results
 - 698 sentences
 - 209 source domain lexical units
 - 134 source frames; 957 frame instances

Case study: Constructions

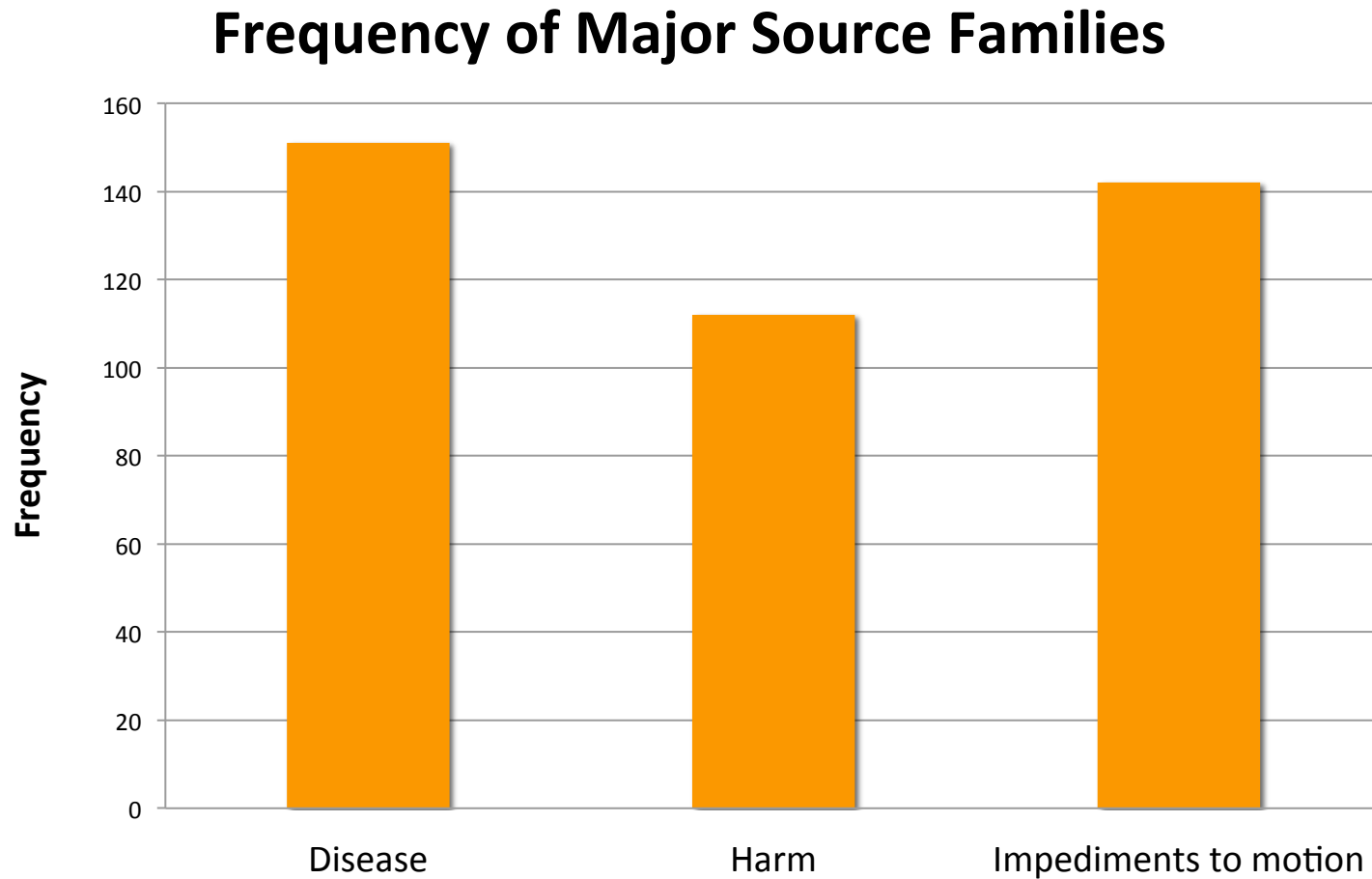


Case study: Source family data

- Highest frequency source domain frames
 - Identified 3 major source frame families
 - Account for 42% of frame instances

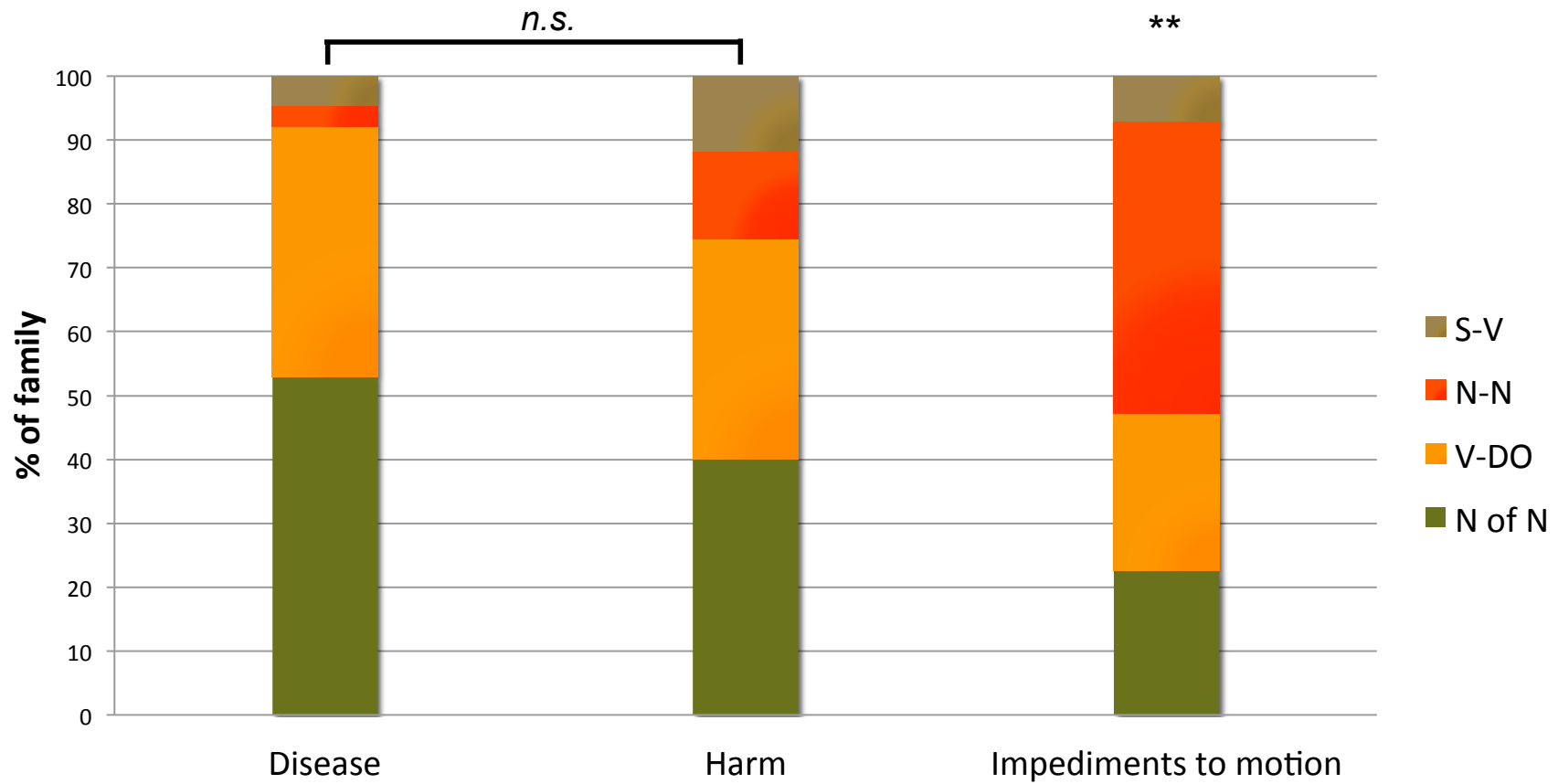
Disease family	Impediments to motion family	Harm family
Diagnosis of affliction	Antagonistic force	Attacking
Disease	Burden	Danger
Health treatment	Constructed confinement	Destroying
Physical affliction	Constructed restraints	Harm
Treating a physical affliction	Remove burden	Harm to living entity
	Sunken confinement	Impact
	Water confinement	Killing

Case study: Major source family data



Constructions in Source family data

Distributions of Constructions by Source Family

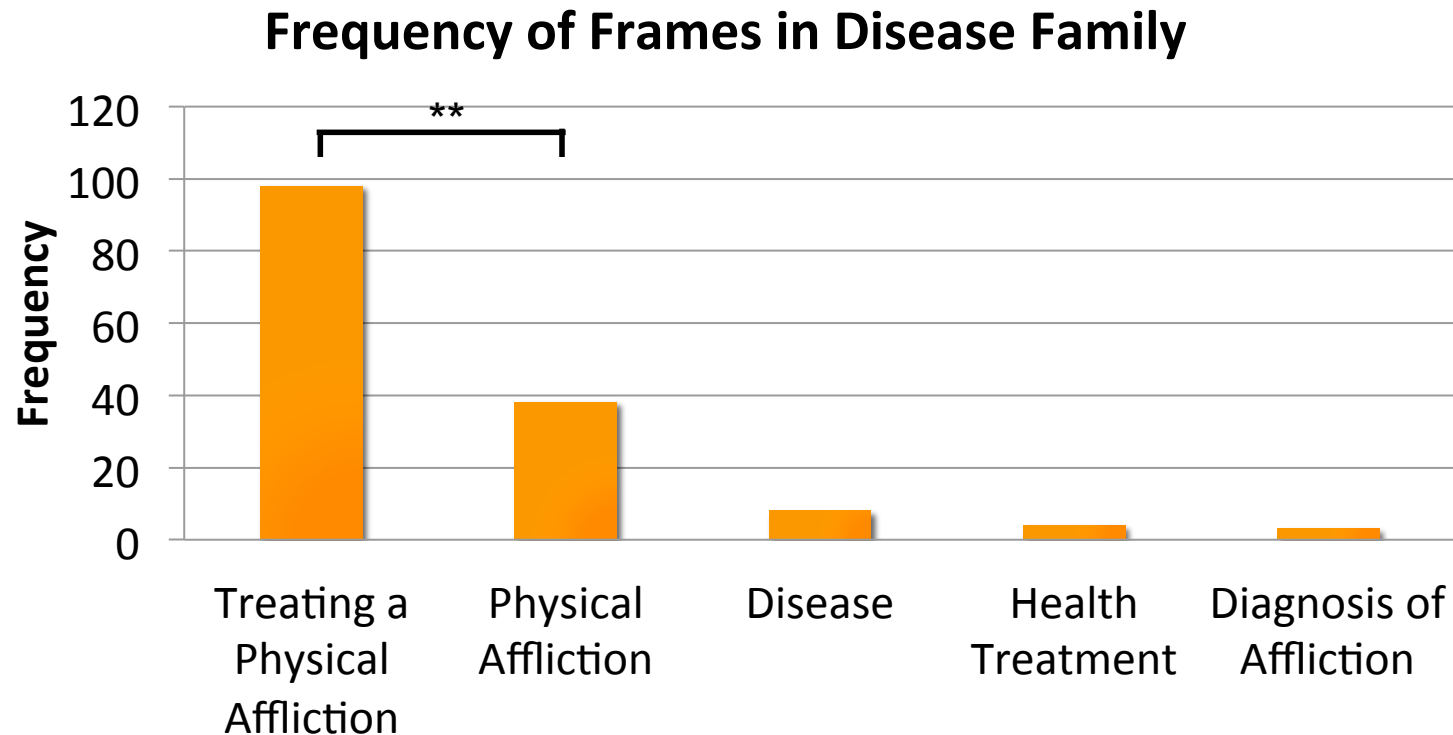


symptoms of poverty
alleviate poverty

impact of poverty
tackle homelessness

poverty trap

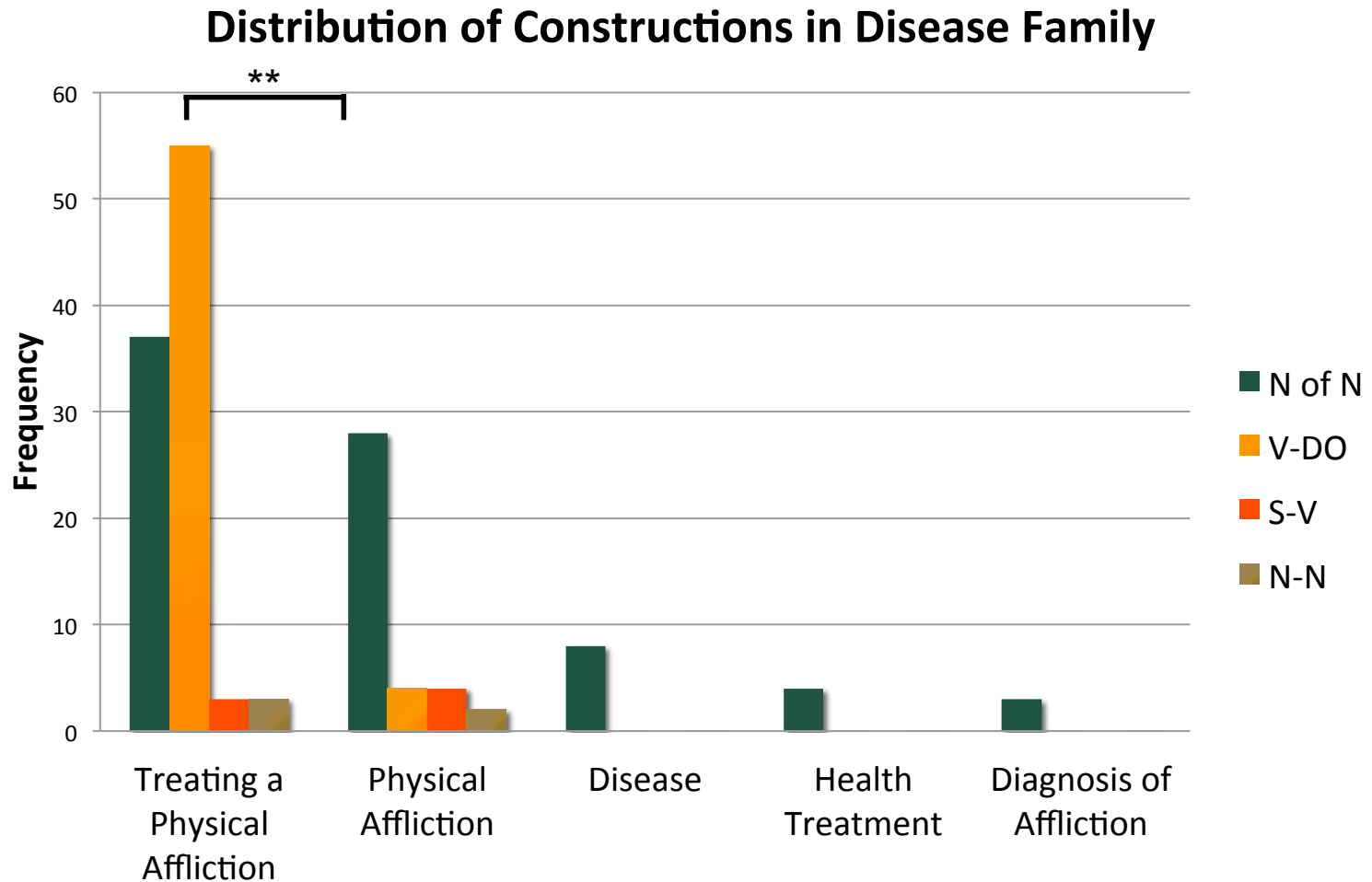
Frames within Disease family



- **Metaphors:**

- POVERTY IS A PHYSICAL AFFLICTION
- ADDRESSING POVERTY IS TREATING A PHYSICAL AFFLICTION

Constructions in Disease family



Constructions in Disease data: Examples

Source frame: Treating a Physical Affliction

Construction: Verb(**Source**)-Direct Object(**Target**)

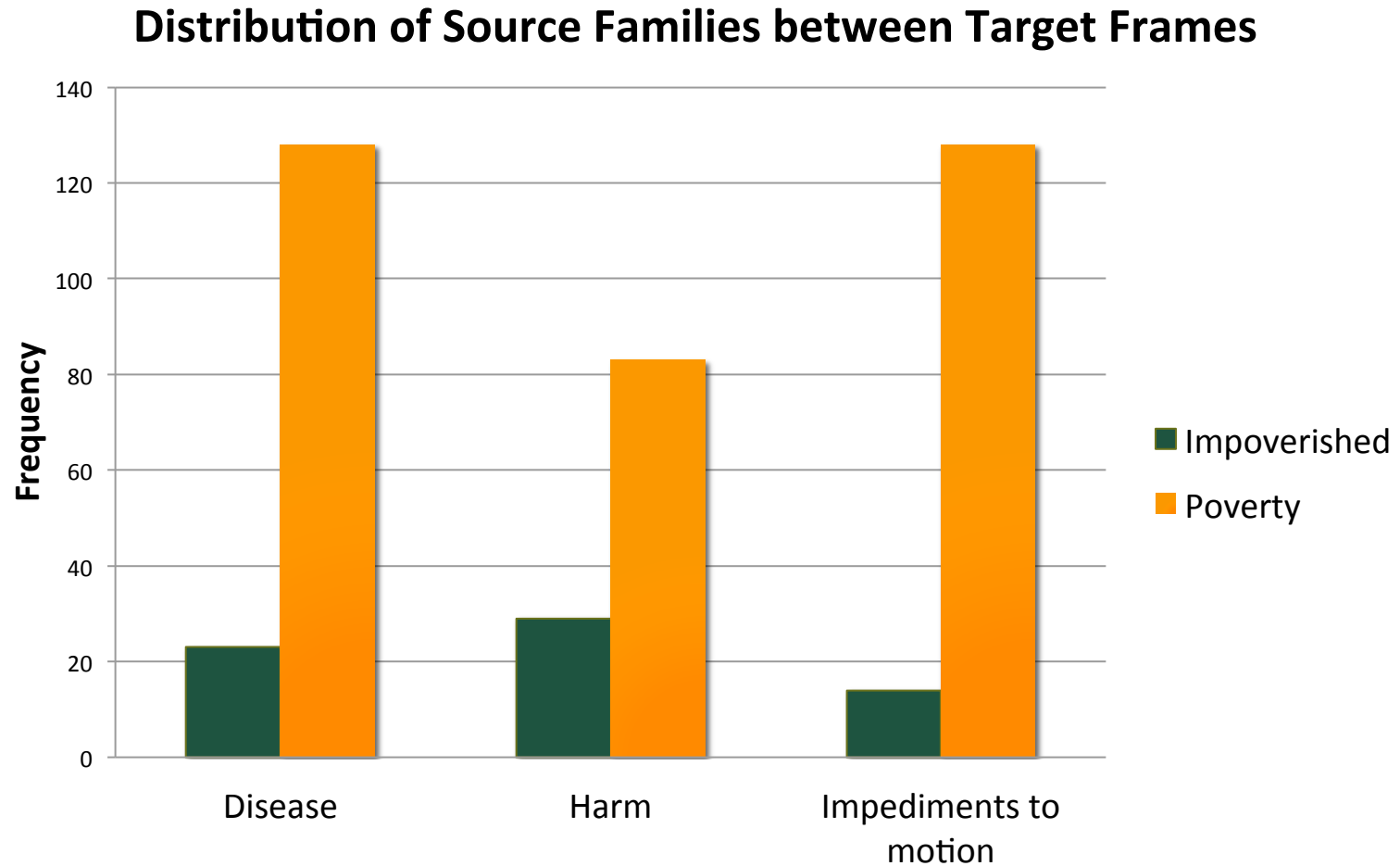
- *palliate serious destitution* (BNC:EE9)
- *eradicate basic poverty* (BNC:A65)
- *alleviate the poverty and suffering of the world* (BNC:ALH)
- *cure deep poverty* (BNC:ABJ)
- *relieve homelessness* (BNC:A4K)

Source frame: Physical affliction

Construction: Noun(**Source**) of Noun(**Target**)

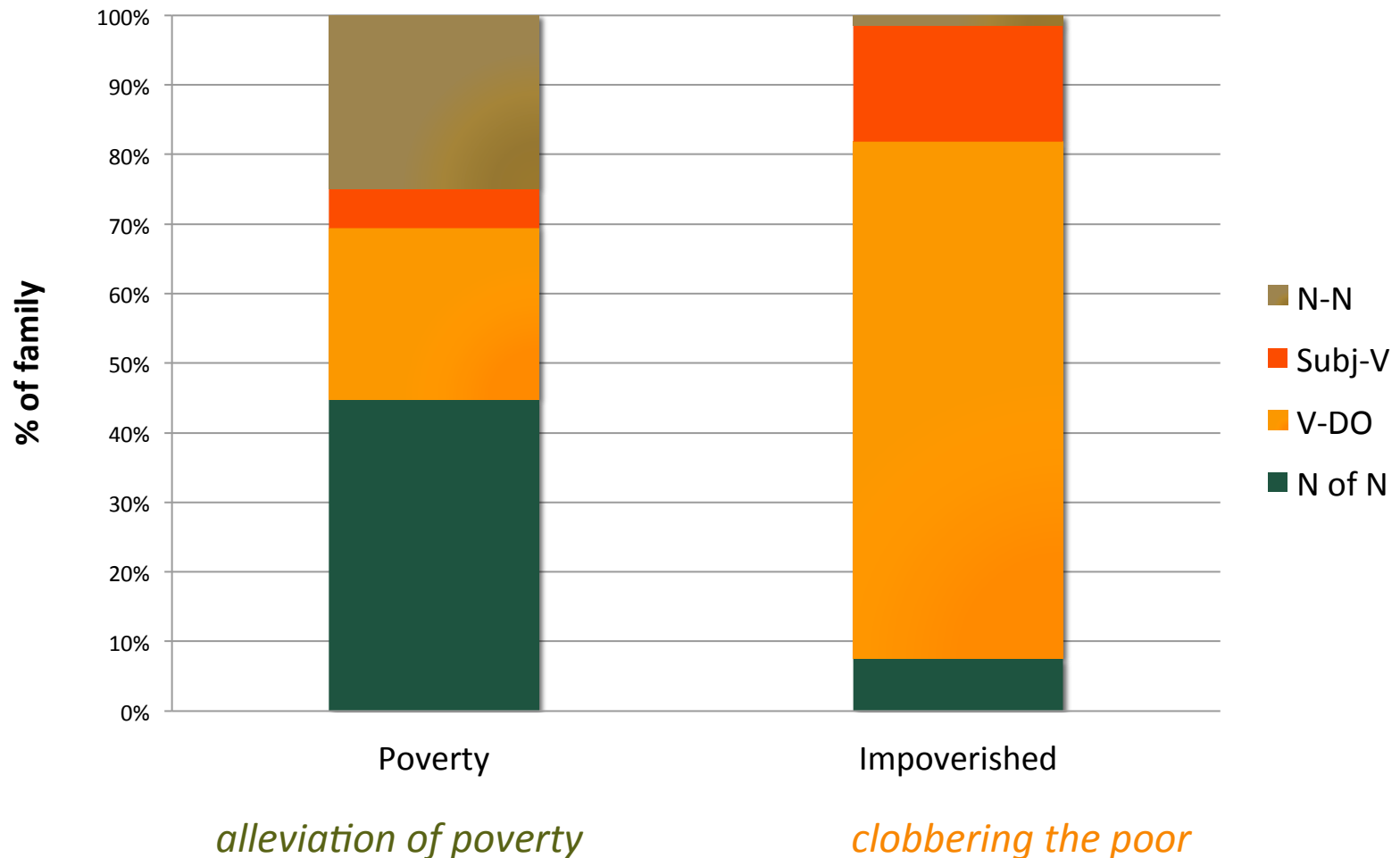
- *the Tory scourge of homelessness* (BNC:HHV)
- *the sickness of poverty* (BNC:ABE)
- *a symptom of poverty* (BNC:APN)

Frames in Target family data



Constructions in Target family data

Distribution of Constructions in Poverty



Conclusions and further progress

- Proof of concept
- Enables rich semantic analysis at different levels of granularity
- Further progress:
 - Defining additional constructions:
 - V(S)-Prep-IO(T), V(S)-Part-Prep-IO(T), Adj-N, N(S)-Prep-N(T), N(S)-Part-Prep-N(S)
 - Continuing to expand the repository's semantic coverage
 - Further testing and refinement of system
 - Cross-linguistic and cross-cultural comparisons
- Initial release of metaphor repository in 2015

References

- *The British National Corpus*, version 3. (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>
- Croft, W. (2002). The role of domains in the interpretation of metaphors and metonymies. *Metaphor and metonymy in comparison and contrast*, 161-205.
- David, O., Dodge, E., Hong, J., Stickles, E., & E. Sweetser. (2014). *Building the MetaNet metaphor repository: The natural symbiosis of metaphor analysis and construction grammar*. Talk presented at the 8th International Conference on Construction Grammar (ICCG 8), Osnabrück, Germany. September 4, 2014.
- Deignan, A. (2005). *Metaphor and corpus linguistics*. John Benjamins.
- Dodge, Ellen; David, Oana; Stickles, Elise, and Eve Sweetser. (2014). *Constructions and metaphor: Integrating MetaNet and Embodied Construction Grammar*. Talk presented at the 8th International Conference on Construction Grammar (ICCG 8), Osnabrück, Germany. September 4, 2014.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. UChicago Press.
- Narayanan, S., & J. Hong. (2013). *A multilingual functional repository for cognitive linguists*. Talk presented at the International Cognitive Linguistics Conference (ICLC) 12, Edmonton, AB. June 24, 2013.
- Shutova, K. (2011). *Computational approaches to figurative language*. PhD thesis, Computer Laboratory, University of Cambridge, UK.
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. John Benjamins.
- Stefanowitsch, A., & Gries, S. T. (Eds.). (2006). *Corpus-based approaches to metaphor and metonymy*. Walter de Gruyter.
- Sullivan, K. (2017). *Grammar in metaphor: A construction grammar account of metaphoric language*. PhD dissertation, UC Berkeley.
- Sullivan, K. (2013). *Frames and constructions in metaphoric language*. John Benjamins.

Acknowledgements

- The MetaNet Analysis and Repository teams: George Lakoff, Eve Sweetser, Oana David, Karie Moorman, Luca Gilardi
- Collin Baker, Jim Hieronymous, and the rest of the MetaNet project members
- Comments and suggestions from Michael Ellsworth
- 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.



Thank you!

- Contact:
 - elstickles@berkeley.edu
 - edodge@icsi.berkeley.edu
 - jhong@icsi.berkeley.edu