Don't 'have a clue'?

Unsupervised co-learning of downward-entailing operators

Cristian Danescu-Niculescu-Mizil, Lillian Lee
Cornell University

Color legend
Target: DE ops. (e.g., doubt)
Clues: NPIs (e.g., any)

Cristi: *"Nicio" ... is that adjective you've mentioned.*
Anca: *A negative pronominal adjective.*
Cristi: *You mean there are people who analyze that kind of thing?*
Anca: *The Romanian Academy.*
Cristi: *They're crazy.*
-- From the movie *Police, adjective*

Lexical semantics phenomenon: Monotonicity

Upward Monotone (default):

*I **know** I'll buy a Mac* ➡ *I **know** I'll buy a computer*
subset of

Downward Monotone:

*I **doubt** I'll buy a Mac* ⬅ *I **doubt** I'll buy a computer*
subset of

Downward entailing operators invert the default monotonicity, allowing one to "reason from sets to subsets" [van den Wouden, 1997]

Examples: (subset of)
He came without cash or cards ➡ *He came without cash*
She is too lazy to run ➡ *She is too lazy to run a 10k*
I am reluctant to steal money ➡ *I am reluctant to steal SEK*

Task: Automatically discover DE operators

Why?

✓ **Linguistic importance**
DE operators play "an extremely important role in natural language".
[van der Wouden, 1997; van Benthem, 1986; Hoeksema, 1986; Dowty, 1994; Sanchez Valencia, 1991]

✓ **Textual entailment**
TE systems rely on small hand-annotated lists of DE operators.
[Nairn et. al, 2006; MacCartney and Manning, 2008; Bar-Haim et al, 2008.]

✓ **Natural language generation**
DM inferences induce greater cognitive load than UM inferences
[Geurts and van der Slik, 2005]

✓ **Prevalence**
At least 6% of newswire sentences contain a non-trivial DE operator
[Danescu-Niculescu-Mizil et al., 2009]

But how?
☹ No monotonicity-annotated corpora
☹ Not deducible from any public lexical database
[Nairn et. al, 2006]

Linguistic insight

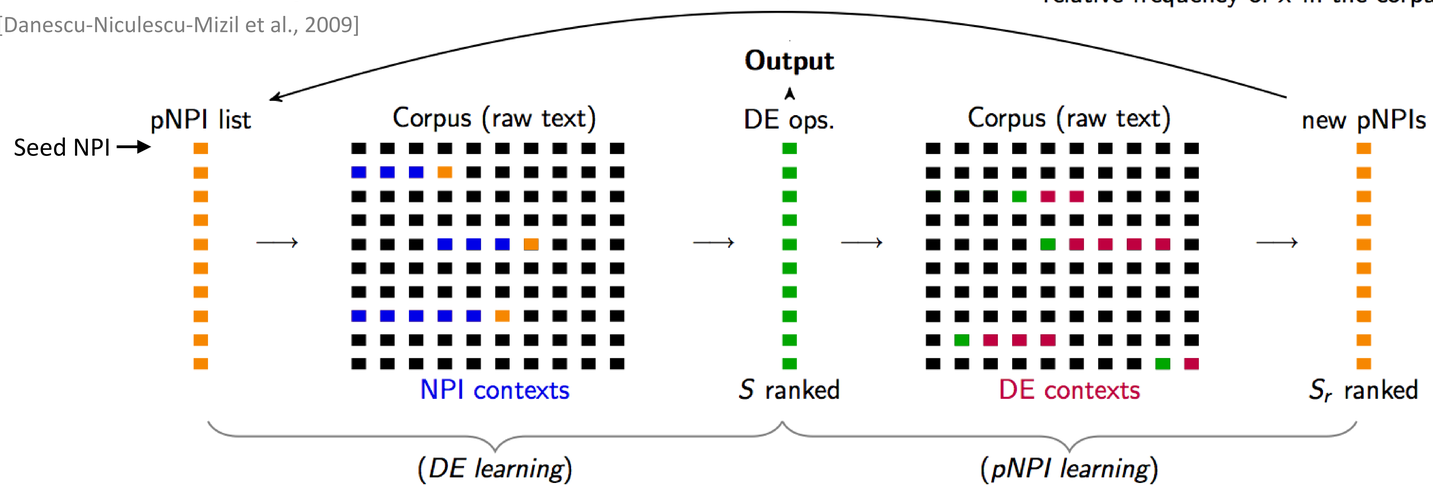
Ladusaw's (1980) Hypothesis:
Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

He does not listen anymore vs. **He does listen anymore*
I doubt you have a clue vs. **You have a clue*

Where's the green?
It is wise to try compensating for any excess.
We have a clue!

$$S(x) := \frac{\text{fraction of NPI contexts that contain } x}{\text{relative frequency of } x \text{ in the corpus}}$$

[Danescu-Niculescu-Mizil et al., 2009]



Clueless!

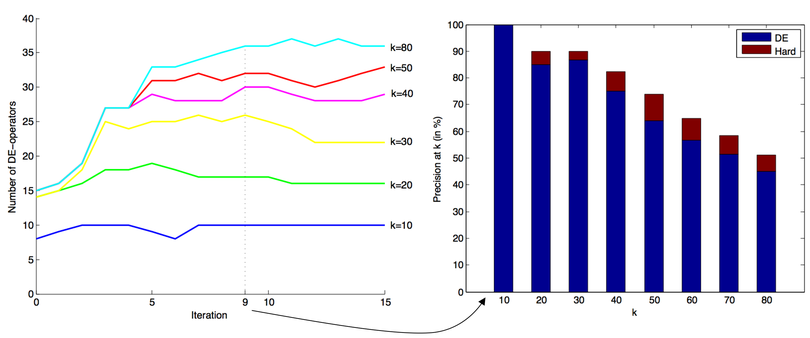
English NPIs	any	have a clue	red cent	give a damn	at all	...
Romanian NPIs	?	?	?	?	?	...
Klingon NPIs	?	?	?	?	?	...

High-quality lists of NPIs are not available for most languages!
☹ NPIs are hard to learn from data. [Lichte and Soehn, 2007; Hoeksema 1997]
☹ Most NPIs do not translate across languages.

Where's the orange?
S-a abţinut să facă vreun comentariu. [this is not Klingon!]

$$S_r(x) := \frac{\text{fraction of DE contexts that contain } x}{\text{relative frequency of } x \text{ in the corpus}}$$

Results for Romanian



First time DE ops. are learned for a language other than English!

System output and annotation available at: www.cs.cornell.edu/~cristian/acl2010

Translation alternative

- Not all DE operators translate.
- 39% of the 36 DE operators discovered in iteration 9 have **no English equivalents** in the largest list of English DE ops available

Does it really work for Klingon?

Q: How to choose the **seed NPI**?
A: Most languages have an NPI that translates to "any".

Q: How strong is the seed?
A: It depends on the language: connection to linguistic typology.
[Haspelmath, 2001]