

Without a 'doubt'?

Unsupervised discovery of downward-entailing operators

Cristian Danescu-Niculescu-Mizil
Lillian Lee
Richard Ducott

Department of Computer Science, Cornell University

06-01-2009 NAACL 2009

What's the deal?

Complex semantic phenomenon, never treated computationally before.

Approach it by using a classic linguistic result.

Unsupervised algorithm achieving precision at k of up to 100%.

What's the deal?

Complex semantic phenomenon, never treated computationally before.

Approach it by using a classic linguistic result.

Unsupervised algorithm achieving precision at k of up to 100%.

Monotonicity

“I know I’ll buy a Mac” \implies “I know I’ll buy a computer”

Monotonicity

“I know I’ll buy a Mac” \implies “I know I’ll buy a computer”

“I doubt I’ll buy a Mac” $\not\Rightarrow$ “I doubt I’ll buy a computer”

Monotonicity

“I know I’ll buy a Mac” \implies “I know I’ll buy a computer”

“I doubt I’ll buy a Mac” \longleftarrow “I doubt I’ll buy a computer”

Monotonicity

Upward monotone:

“I know I’ll buy a Mac” \Rightarrow “I know I’ll buy a computer”
<subset> <set>

Downward monotone:

“I doubt I’ll buy a Mac” \Leftarrow “I doubt I’ll buy a computer”
<subset> <set>

Monotonicity

Upward monotone:

“I know I’ll buy a Mac” \Rightarrow “I know I’ll buy a computer”
<subset> <set>

Downward monotone:

“I **doubt** I’ll buy a Mac” \Leftarrow “I **doubt** I’ll buy a computer”
<subset> <set>

Downward-entailing operators

Definition:

Downward-entailing operators invert the default monotonicity, allowing one to “reason from sets to subsets”

[van der Wouden, 1997]

Downward-entailing operators

Definition:

Downward-entailing operators invert the default monotonicity, allowing one to “reason from sets to subsets”

[van der Wouden, 1997]

Examples:

“I **doubt** I’ll buy a computer” \implies “I **doubt** I’ll buy a Mac”

“He came **without** cash or cards” \implies “He came **without** cash”

“She was **too lazy to** run” \implies “She was **too lazy to** run a 10k”

Downward-entailing operators

Task:

Automatically discover downward-entailing operators

Downward-entailing operators

Why?

Linguistic importance:

DE operators play “an extremely important role in natural language”
[van der Wouden, 1997; van Benthem, 1986; Hoekesema, 1986;
Dowty, 1994; Sánchez Valencia, 1991]

Textual Entailment:

Textual Entailment systems that approach monotonicity rely on relatively small hand-annotated lists of DE operators.
[Nairn et al.; 2006, MacCartney and Manning, 2008;
Bar-Haim et al., 2008.]

Prevalence:

Estimate (post-hoc) that at least 6% of newswire sentences contain a non-trivial DE operator (excluded: ‘not’, ‘no’, ‘none’, ‘few’, etc.).

Downward-entailing operators

Why?

Linguistic importance:

DE operators play “an extremely important role in natural language”
[van der Wouden, 1997; van Benthem, 1986; Hoekesema, 1986;
Dowty, 1994; Sánchez Valencia, 1991]

Textual Entailment:

Textual Entailment systems that approach monotonicity rely on relatively small **hand-annotated** lists of DE operators.
[Nairn et al.; 2006, MacCartney and Manning, 2008;
Bar-Haim et al., 2008.] → **only about 20 DE ops.**

Prevalence:

Estimate (post-hoc) that at least 6% of newswire sentences contain a non-trivial DE operator (excluded: ‘not’, ‘no’, ‘none’, ‘few’, etc.).

Downward-entailing operators

Why?

Linguistic importance:

DE operators play “an extremely important role in natural language”
[van der Wouden, 1997; van Benthem, 1986; Hoekesema, 1986;
Dowty, 1994; Sánchez Valencia, 1991]

Textual Entailment:

Textual Entailment systems that approach monotonicity rely on relatively small hand-annotated lists of DE operators.
[Nairn et al.; 2006, MacCartney and Manning, 2008;
Bar-Haim et al., 2008.]

Prevalence:

Estimate (post-hoc) that at least 6% of newswire sentences contain a non-trivial DE operator (excluded: ‘not’, ‘no’, ‘none’, ‘few’, etc.).

Downward-entailing operators

Challenges

No monotonicity-annotated corpora.

Not deducible from any public lexical database. [Nairn et al., 2006]

Downward-entailing operators

Challenges

No monotonicity-annotated corpora.

Not deducible from any public lexical database. [Nairn et al., 2006]

We need something else to hang on to!

Luck?

Downward-entailing operators

Challenges

No monotonicity-annotated corpora.

Not deducible from any public lexical database. [Nairn et al., 2006]

We need something else to hang on to!

Luck? → free “noisy annotation”

Remember our deal?

Complex semantic phenomenon, never treated computationally before.

Approach it by using a classic linguistic result.

Unsupervised algorithm achieving precision at k of up to 100%.

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Definition:

Negative Polarity Items (NPIs) are terms that tend to occur in “negative environments”.

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Examples:

“They do not listen **anymore.**” vs. “*They do listen **anymore.**”

“I doubt they **give a damn.**” vs. “*They **give a damn.**”

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Examples:

“They do **not** listen **anymore**.” vs. “*They do listen **anymore**.”

“I **doubt** they **give a damn**.” vs. “*They **give a damn**.”

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Examples:

“They do **not** listen **anymore**.” vs. “*They do listen **anymore**.”

“I **doubt** they **give a damn**.” vs. “*They **give a damn**.”

Note:

This hypothesis is heavily debated in the linguistic literature.

Negative Polarity Items

Ladusaw's (1980) Hypothesis:

Negative Polarity Items (NPIs) only appear within the scope of downward-entailing operators.

Examples:

“They do **not** listen **anymore.**” vs. “*They do listen **anymore.**”

“I **doubt** they **give a damn.**” vs. “*They **give a damn.**”

Where's the green?

“It is wise to try compensating for **any** excess.”

NPI context

The deal...

Complex semantic phenomenon, never treated computationally before.

Approach it by using a classic linguistic result.

Unsupervised algorithm achieving precision at k of up to 100%.

Distributional assumption

Ladusaw's Hypothesis → Distributional Assumption

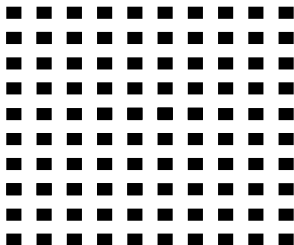
DE operators have a higher frequency in **NPI contexts** than in the whole corpus.

Distributional Assumption → Candidate Score

$$\text{Candidate Score} = \frac{\text{Freq. in NPI contexts}}{\text{Freq. in Corpus}}$$

Distributional assumption

Corpus (raw text)

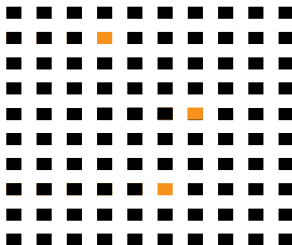


Distributional assumption

NPI list



Corpus (raw text)

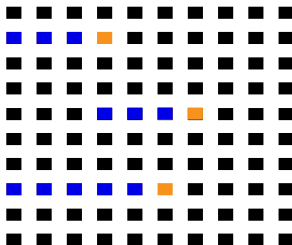


Distributional assumption

NPI list

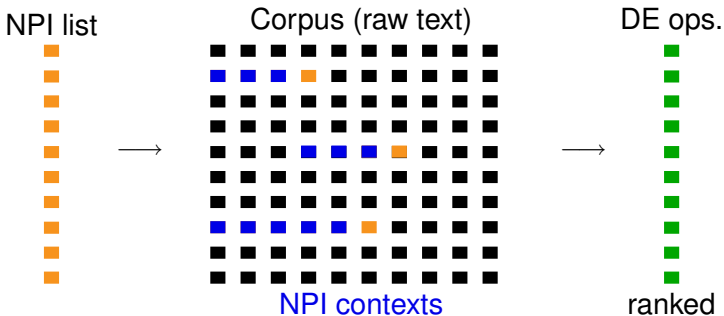


Corpus (raw text)



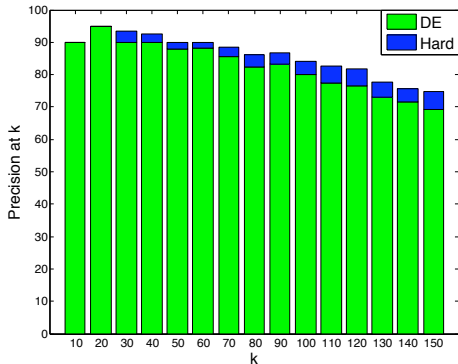
NPI contexts

Distributional assumption



Results

Precision at k , for $k = 10$ to 150



Data

BLLIP 1987-89 (raw text only)

1,796,379 sentences

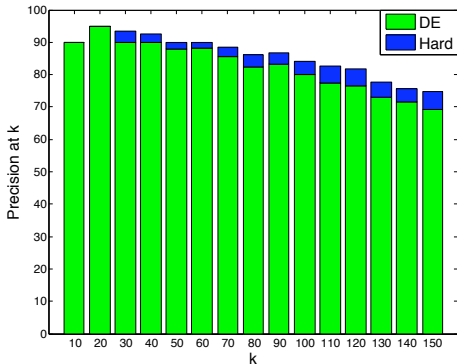
List of NPIs from Lawler (2005)

26 NPIs

53,064 NPI contexts

Results

Precision at k , for $k = 10$ to 150



Recall

Can not be evaluated directly, because no comprehensive list of DE operators exists.

We retrieve almost all of the (about 20) hand-annotated operators.

Distillation

Piggybackers

“He vigorously denies any accusations.”

“This was biggest one-day drop in years.”

Piggybackers occur frequently with DE ops.,
and rarely without.

→ Undeserved high score

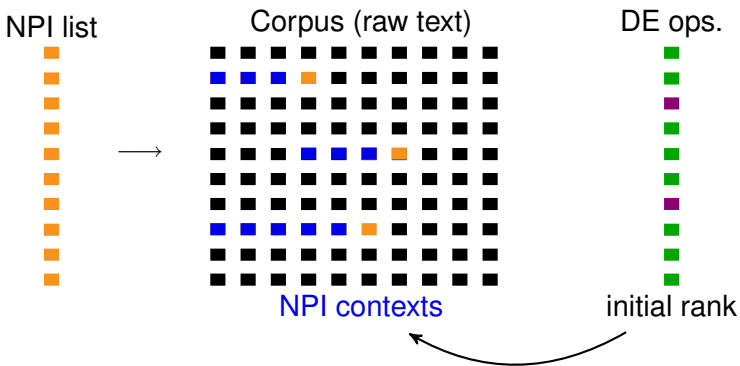
Solution: Make the terms in an NPI context compete
for its “budget” using their initial relative scores.
(see paper for details)

DE ops.

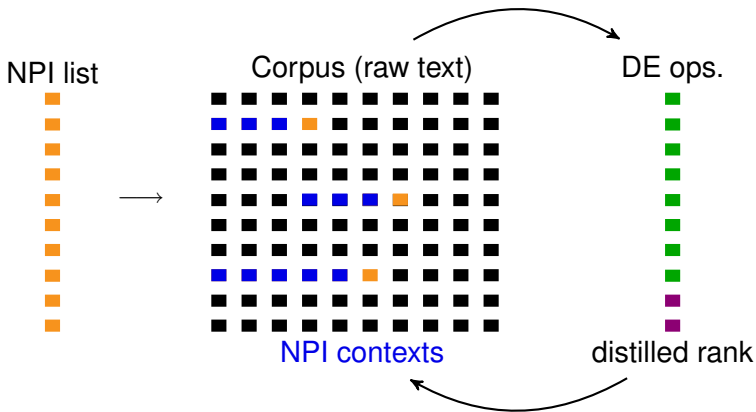


ranked

Distillation

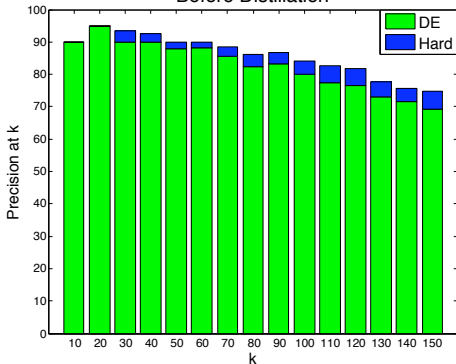


Distillation

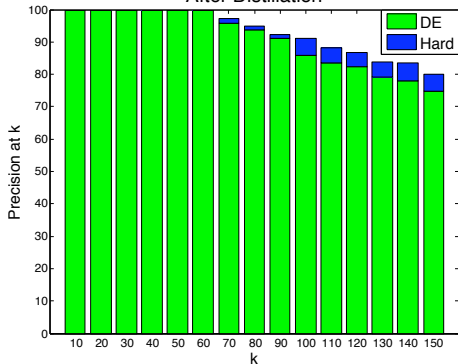


Results

Before Distillation



After Distillation



Related Work

Textual Entailment

[MacCartney and Manning, 2008]:

Integrate monotonicity in an entailment system.

Use a relatively small hand-crafted lists of **DE operators**.

Acknowledge the importance of **DE operators** for entailment.

[Nairn et al.; 2006, Bar-Haim et al., 2008]:

“Polarity” in textual entailment (related concept).

Also make use of a hand-crafted list containing **DE operators**.

Linguistic Literature

The focus is on **NPIs**, rather than on **DE operators**.

[Lichte and Soehn, 2007]: Corpus based discovery of German **NPIs**.

[Hoeksema, 1997]: Acknowledges the difficulty of finding **DE ops**.

Conclusions

Complex semantic phenomenon, never treated computationally before.

Approach it by using a classic linguistic result.

Unsupervised algorithm achieving precision at k of up to 100%.

Conclusions

Complex semantic phenomenon, never treated computationally before:
→ automatic discovery of downward-entailing operators.

Approach it by using a classic linguistic result:
→ Ladusaw's contested hypothesis relating DE operators with NPIs.

Unsupervised algorithm achieving precision at k of up to 100%:
→ discovered many novel and non-obvious DE operators
→ resource lean: needs only raw text and a list of NPIs
→ works also for languages where list of NPIs are not available
(current work)

Thank You!

Complex semantic phenomenon, never treated computationally before:
→ automatic discovery of downward-entailing operators.

Approach it by using a classic linguistic result:
→ Ladusaw's contested hypothesis relating DE operators with NPIs.

Unsupervised algorithm achieving precision at k of up to 100%:
→ discovered many novel and non-obvious DE operators
→ resource lean: needs only raw text and a list of NPIs
→ works also for languages where list of NPIs are not available
(current work)