Hedge Detection as a Lens on Framing in the GMO Debates

A Position Paper

Eunsol Choi*, Chenhao Tan*, Lillian Lee*,
Cristian Danescu-Niculescu-Mizil* and Jennifer Spindel[+]

**Computer Science*, Plant Breeding and Genetics[+]
Cornell University**

1

Keywords

Framing

Scientific Discourse

Hedging

Framing

- Can we learn how people are framing the issue from their use of language?

  Small changes in the presentation of an issue produce changes of opinion [Chong and Druckman (2007)]

  e.g. "pro-life" VS "pro-choice", "freedom fighters" VS "terrorist"

Framing in GMO debates

- In this paper: Do proponents and opponents of genetically modified organisms (GMO) use different language?

Framing in GMO debates

- Lexical Choice: word choice in synonyms
  - "Transgenic food" VS "Frankenfood"

- Stylistic Difference:
  - Pro-GMO people are a scientific crowd,
    while anti-GMO people are laypeople?
  - What does it mean for a text to be "scientific"?

"Scientific" Discourse

Question from rhetoric and communication studies:

Do professional science articles include more uncertainty than popular science ones?

"Scientific" Discourse   Hedging

Do professional science articles include more uncertainty than popular science ones?

Hedging (Hyland, 1998)

– an expression of tentativeness and possibility

– the writer withholding full commitment to statements

Do professional science articles include more **uncertainty** than popular science ones?

Hedging (Hyland, 1998)

example

1) It _seems that_ this group plays a critical role

2) We *wish* _to suggest_ a structure for the salt of deoxyribose nucleic acid (DNA).

More Scientific = More Hedging

- Hedges are abundant in science and play a critical role in academic writing. (Hyland 1998)

- "The hedging of claims is so common that a sentence that looks like a claim but has no hedging is probably not a statement of new knowledge" (Myers 1989)

- Scientific observations to popular accounts brings "removing hedges … thus conferring greater certainty on the reported facts" (Fahnestock 1986)
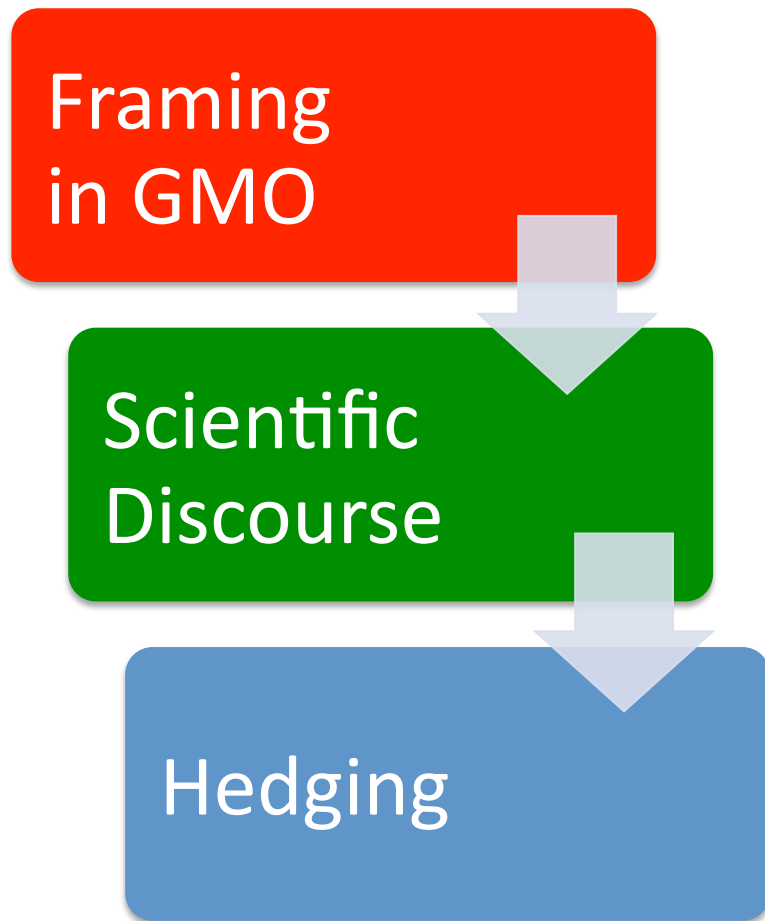
More Scientific = More Hedging ??

No Consensus

- Hedging is "typical of popular scientific articles" (Varttala 1999)

Prior research was...

less computational, small-scale analyses

Research Question

Framing in GMO

Scientific Discourse

Hedging

Is hedging, considered as a single feature, correlated with scientific discourse or being a proponent/ opponent of GMOs.

Procedures Sketch

This Paper: Investigate relative degree of hedging in scientific VS. non-scientific documents

More Scientific	= More Hedging ?

Procedures Sketch

This Paper: Investigate relative degree of hedging in scientific VS. non-scientific documents

More Scientific	= More Hedging ?

Future Work: Investigate whether the use of hedging in pro-GMO articles follows our inferred "scientific" occurrence patterns.

Framing in GMO	Scientific Discourse	Hedging

2010 CoNLL Shared Task:
Identifying hedges and their scopes

- Learning to detect sentences containing uncertainty and its scope

- Providing BioScope and Wikipedia weasel annotation data with cue words and weasel words

2010 CoNLL Shared Task:
Identifying hedges and their scopes

- Learning to detect sentences containing uncertainty and its scope

- Providing BioScope and Wikipedia weasel annotation data with cue words and weasel words

e.g.)
1) There is a $20 per automobile user fee to enter the park.-------------------- Certain
2) Children *appear* to rarely make up false allegations of their own accord.-- Uncertain
3) *Many accounts*, including *Muslim accounts*, and *some accounts written* -- Uncertain
by academic historians, stress the power and importance of the pre-Islamic
Mecca

2010 CoNLL Shared Task:
Identifying hedges and their scopes

- Learning to detect sentences containing uncertainty and its scope

- Providing BioScope and Wikipedia weasel annotation data with cue words and weasel words

e.g.)
1) There is a $20 per automobile user fee to enter the park.-------------------- Certain
2) Children *appear* to rarely make up false allegations of their own accord.-- Uncertain
3) *Many accounts*, including *Muslim accounts*, and *some accounts written* -- Uncertain
by academic historians, stress the power and importance of the pre-Islamic
Mecca

Weasel word

Reimplementation of CoNLL winner

- Best-Performing algorithm in Wikipedia Hedge detection task : **Georgescul**

- SVM classification based on a Gaussian Radial Basis kernel function

- Features: the frequency of cue word and its component 2grams and 3grams from the training corpus

  Used LIBSVM and tuned parameters through cross validation

Data set

GMO Data Set

Pop-Sci (648 docs)		Pro-GMO (671 docs)
Prof-Sci (928 docs)	Pop-Sci / Prof-Sci 200 annotated hedge sentences	Anti-GMO (762 docs)

Pop-Sci: US newspapers articles from LexisNexis
Prof-Sci: scientific journal and conference abstracts from Web of Science

Pro and anti-GMO articles were collected from strongly opinionated blogs

Annotation Effort

GMO Annotation Set:

- – Randomly chosen sentences 100 from Pop-Sci and 100 from Prof-Sci

- – Cohen's Kappa : 0.67 → fair or good consistency

Hedge Annotation is not trivial!

- – Cassava is the staple food of tropical Africa and its production, averaged over 24 countries, has increased *more than* threefold from 1980 to 2005 …

Data set

GMO Data Set

Pop-Sci (648 docs)

Prof-Sci (928 docs)

Pop-Sci / Prof-Sci 200 annotated hedge sentences

Pro-GMO (671 docs)

Anti-GMO (762 docs)

Training Data Set from CoNLL Shared Task

Wikipedia set (20745 sentences)

BioScope set (19544 sentences)

Set of Hedge Cues

Hedge classifier
In-domain evaluation

5 fold Cross-Validation using CoNLL dataset

Dataset	Precision	Recall	F-Score
Bio	84.0	92.0	87.8
Wiki	64.0	76.3	69.6
Bio + Wiki	66.7	78.3	72.0

"Unsatisfying"
Cross-domain evaluation

Trained on CoNLL data, tested on GMO data

Test data	Training data	Precision	Recall	F-Score
Prof-Sci+Pop-Sci	Bio	54	68	60
Prof-Sci+Pop-Sci	Wiki	38	54	45
Prof-Sci+Pop-Sci	Bio+Wiki	21	93	34

Why "unsatisfying" result?

Simplicity of algorithm

- Longer sentences tend to be classified as uncertain

- Misleading bigrams and trigrams

  "certain leisure or cultural events", "people of Jewish tradition", "some of schumann's best choral writing"

Feature Source	# Features
Bio / Bio (cue+2gram+3gram)	220 / 340
Wiki / Wiki (cue+2gram+3gram)	3740 / 10603

Why "unsatisfying" result?

Domain difference between training and test data

- Bio model performed better on Prof-Sci than in Pop-Sci

	Precision	Recall	F-Score
Prof-Sci	58	73	65
Pop-Sci	52	64	57

- Wiki model performed better on Pop-Sci than in Prof-Sci.

% of Uncertain Sentences

Dataset	% Uncertain Sentences
Annotated Prof-Sci	20
Annotated Pop-Sci	28
BIO	17*
WIKI	23*

SVM Classifier from our study

Dataset	Model	% Uncertain Sentences
Prof-Sci	Bio	16*
Pop-Sci	Bio	19*

Hedging *seems to* appear more frequently in popular science articles.

Conclusion

Correlation between hedging and scientific discourse still remains to be discovered

– Need a more reliable, general hedge classifier

$\rightarrow$ **"A Position Paper"**

What are we positioning?

First Step: Investigate relative degree of hedging in scientific VS. non-scientific documents

More Scientific	= More Hedging ?

Second Step: Investigate whether the use of hedging in pro-GMO articles follows our inferred "scientific" occurrence patterns.

Framing in GMO	Scientific Discourse	Hedging

Contribution

- Another Perspective to view "Hedging"
  - In terms of its implication in framing
- Constructed and released four GMO-related, distinct context datasets

(https://confluence.cornell.edu/display/llresearch/HedgingFramingGMOs)

Thank You

* this is a certain statement!