Studying the Distribution of Fragments of English Using Deep Semantic Annotation

Camilo Thorne[1]

[1] KRDB Research Centre for Knowledge and Data
Free University of Bozen-Bolzano
http://www.inf.unibz.it/~cathorne
cthorne@inf.unibz.it

ISA-8 Workshop, Pisa, 4/10/2012

Outline

Motivation

- Suppose we came across the following argument in some text

  Every Italian loves pasta and football
  Silvio is Italian

  that entails

  Silvio loves pasta

Motivation

- Suppose we came across the following argument in some text

  Every Italian loves pasta and football
  Silvio is Italian

  that entails

  Silvio loves pasta

- Common-sense reasoning

- How difficult can it be to reason over such arguments?

Motivation

- Natural logic studies the complexity of common-sense reasoning in natural language [Mus10, Mos10]

Motivation

- Natural logic studies the complexity of common-sense reasoning in natural language [Mus10, Mos10]
- One technique $\Rightarrow$ computational complexity of formal semantics

Motivation

- Natural logic studies the complexity of common-sense reasoning in natural language [Mus10, Mos10]
- One technique $\Rightarrow$ computational complexity of formal semantics
- Known as semantic complexity
- Studied by logicians and semanticists alike
- How often constructs with high semantic complexity occur in corpora?

Motivation

- Natural logic studies the complexity of common-sense reasoning in natural language [Mus10, Mos10]
- One technique ⇒ computational complexity of formal semantics
- Known as semantic complexity
- Studied by logicians and semanticists alike
- How often constructs with high semantic complexity occur in corpora?

⇒ Can we leverage on state-of-the-art deep semantic annotation?

Motivation

- Natural logic studies the complexity of common-sense reasoning in natural language [Mus10, Mos10]
- One technique ⇒ computational complexity of formal semantics
- Known as semantic complexity
- Studied by logicians and semanticists alike
- How often constructs with high semantic complexity occur in corpora?

⇒ Can we leverage on state-of-the-art deep semantic annotation?

QUE: Is semantic complexity inversely proportional to frequency?

The Fragments of English [PHT06]

- The fragments of English are linguistically salient, ambiguity-free subsets of English [PHT06]

The Fragments of English [PHT06]

- The fragments of English are linguistically salient, ambiguity-free subsets of English [PHT06]

1. polynomially translate to FO meaning representations φ

The Fragments of English [PHT06]

- The fragments of English are linguistically salient, ambiguity-free subsets of English [PHT06]

1. polynomially translate to Fo meaning representations φ
2. basic machinery of formal semantics [Mon70]

The Fragments of English [PHT06]

- The fragments of English are linguistically salient, ambiguity-free subsets of English [PHT06]

1. polynomially translate to Fo meaning representations φ
2. basic machinery of formal semantics [Mon70]

- Defined by constraining syntax, semantics and vocabulary
- Used to provide lightweight English front-ends to OWL ontologies [FK06]

The Fragments of English [PHT06]

- The fragments of English are linguistically salient, ambiguity-free subsets of English [PHT06]

1. polynomially translate to Fo meaning representations φ
2. basic machinery of formal semantics [Mon70]

- Defined by constraining syntax, semantics and vocabulary
- Used to provide lightweight English front-ends to OWL ontologies [FK06]

$\Rightarrow$ Efficient translation, but how costly is logical reasoning?

The Fragments of English [PHT06, Tho10]

Fragment	Coverage	Fo Operators
COP($\neg$)	Copula ("is a"), nouns ("man"), intransitive verbs ("runs"), "every", "some" names ("Joe"), adjectives ("thin") (+"not")	$\{\forall, \exists, (\neg)\}$
COP($\neg$)+TV	COP($\neg$) +transitive verbs ("loves")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+DTV	COP($\neg$) +ditransitive verbs ("gives")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+TV +DTV	COP($\neg$)+TV + ditransitive verbs	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+Rel	COP($\neg$)+ ("who", "that", "which") "and", intersective adjectives (+"or")	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV	COP($\neg$)+Rel +transitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +DTV	COP($\neg$)+Rel +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV+DTV	COP($\neg$)+Rel+TV +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$

The Fragments of English [PHT06, Tho10]

Fragment	Coverage	Fo Operators
COP($\neg$)	Copula ("is a"), nouns ("man"), intransitive verbs ("runs"), "every", "some" names ("Joe"), adjectives ("thin") (+"not")	$\{\forall, \exists, (\neg)\}$
COP($\neg$)+TV	COP($\neg$) +transitive verbs ("loves")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+DTV	COP($\neg$) +ditransitive verbs ("gives")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+TV +DTV	COP($\neg$)+TV + ditransitive verbs	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+Rel	COP($\neg$)+ ("who", "that", "which") "and", intersective adjectives (+"or")	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV	COP($\neg$)+Rel +transitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +DTV	COP($\neg$)+Rel +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV+DTV	COP($\neg$)+Rel+TV +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$

The Fragments of English [PHT06, Tho10]

Fragment	Coverage	Fo Operators
COP($\neg$)	Copula ("is a"), nouns ("man"), intransitive verbs ("runs"), "every", "some" names ("Joe"), adjectives ("thin") (+"not")	$\{\forall, \exists, (\neg)\}$
COP($\neg$)+TV	COP($\neg$) +transitive verbs ("loves")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+DTV	COP($\neg$) +ditransitive verbs ("gives")	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+TV +DTV	COP($\neg$)+TV + ditransitive verbs	$\{\forall, \exists, (\neg)\}\cup$
COP($\neg$)+Rel	COP($\neg$)+ ("who", "that", "which") "and", intersective adjectives (+"or")	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV	COP($\neg$)+Rel +transitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +DTV	COP($\neg$)+Rel +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$
COP($\neg$)+Rel +TV+DTV	COP($\neg$)+Rel+TV +ditransitive verbs	$\{\forall, \exists, \wedge, (\neg, \vee)\}$

The Fragments of English (Examples)

Fragment	**Example**	**Fo**
COP	Every politician cheats	$\forall x(\mathit{Politician}(x) \rightarrow \mathit{Cheat}(x))$
COP$^{\neg}$	Some philosopher is not trustworthy	$\exists x(\mathit{Philosopher}(x) \wedge \neg \mathit{Trusted}(x))$
COP$^{\neg}$+TV	John does not love Luke	$\neg \mathit{Loves}(\text{John}, \text{Luke})$
COP+TV +DTV	John gives a book to Jane Some man likes every candy	$\exists x \mathit{Book}(x) \wedge$ $\mathit{Gives}(\text{John}, x, \text{Jane})$ $\exists x(\mathit{Man}(x) \wedge$ $\forall y \mathit{Candy}(x) \rightarrow \mathit{Likes}(x, y))$
COP +Rel	Every idiot who is a philosopher cheats	$\forall x(\mathit{Idiot}(x) \wedge \mathit{Philosopher}(x)$ $\rightarrow \mathit{Cheat}(x))$
COP$^{\neg}$ +Rel	Some man who does not cheat is trustworthy	$\forall x(\mathit{Man}(x) \wedge \neg \mathit{Cheat}(x)$ $\rightarrow \mathit{Trusted}(x))$
$\vdots$	$\vdots$	$\vdots$

Semantic Complexity [PHT06, Tho10]

- The fragments of English give rise to semantic complexity
- Defined as the computational complexity of reasoning with of Fo meaning representations

Semantic Complexity [PHT06, Tho10]

- The fragments of English give rise to semantic complexity
- Defined as the computational complexity of reasoning with of FO meaning representations
- It turns out [PHT06, Tho10] that, in general
  1. fragments that contain either negation or relatives, but not both have tractable (polynomial) complexity
  2. fragments that cover both negation and relatives, but not both have intractable (exponential) complexity $\Rightarrow$ encode Boolean satisfiability

Semantic Complexity [PHT06, Tho10]

- The fragments of English give rise to semantic complexity
- Defined as the computational complexity of reasoning with of Fo meaning representations
- It turns out [PHT06, Tho10] that, in general
  1. fragments that contain either negation or relatives, but not both have tractable (polynomial) complexity
  2. fragments that cover both negation and relatives, but not both have intractable (exponential) complexity $\Rightarrow$ encode Boolean satisfiability

$\Rightarrow$ People have a hard time processing intractable English constructs [Szy09]

$\Rightarrow$ What about corpora?

Semantic Annotation - Boxer [Bos08] & Patterns

1. Exploit deep semantic parsers, in particular Boxer 2.0 [Bos08]

Semantic Annotation - Boxer [Bos08] & Patterns

1. Exploit deep semantic parsers, in particular Boxer 2.0 [Bos08]
   - When parsing Wh-questions from the TREC 2008

     What is one common element of major religions?

     Boxer outputs

$$\exists y \exists z \exists e \exists u(\texttt{card}(y,u) \wedge \texttt{c1num}(u) \wedge \texttt{nnumeral1}(u) \wedge \texttt{acommon1}(y) \wedge \texttt{nelement1}(y) \wedge \texttt{amajor1}(z) \wedge \texttt{nreligions1}(z) \wedge \texttt{nevent1}(e) \wedge \texttt{rof1}(y,z))$$

Semantic Annotation - Boxer [Bos08] & Patterns

1. Exploit deep semantic parsers, in particular Boxer 2.0 [Bos08]
   - When parsing Wh-questions from the TREC 2008

     What is one common element of major religions?

     Boxer outputs

     $$\exists y \exists z \exists e \exists u(\texttt{card}(y,u) \wedge \texttt{c1num}(u) \wedge \texttt{nnumeral1}(u) \wedge \texttt{acommon1}(y) \wedge \texttt{nelement1}(y) \wedge \texttt{amajor1}(z) \wedge \texttt{nreligions1}(z) \wedge \texttt{nevent1}(e) \wedge \texttt{rof1}(y,z))$$

   - $\wedge$ and $\exists$ co-occur, but not $\vee$, $\neg$, or $\forall$

Semantic Annotation - Boxer [Bos08] & Patterns

1. Exploit deep semantic parsers, in particular Boxer 2.0 [Bos08]
   - When parsing Wh-questions from the TREC 2008

     What is one common element of major religions?

     Boxer outputs

     $$\exists y \exists z \exists e \exists u(\texttt{card}(y, u) \wedge \texttt{c1num}(u) \wedge \texttt{nnumeral1}(u) \wedge \texttt{acommon1}(y) \wedge \texttt{nelement1}(y) \wedge \texttt{amajor1}(z) \wedge \texttt{nreligions1}(z) \wedge \texttt{nevent1}(e) \wedge \texttt{rof1}(y, z))$$

   - $\wedge$ and $\exists$ co-occur, but not $\vee$, $\neg$, or $\forall$

2. Does Boxer beat patterns?
   - for $\neg$: "not", "no"
   - for $\exists$: "some", "a"
   - for $\forall$: "all", "every, "each"
   - for $\wedge$: "who", "what", "which", "and"
   - for $\vee$: "or"

Classifying Sentences

1. Annotate the corpora/mine the patterns

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

(a) 4 "non-Boolean-closed" classes: $\{\exists, \wedge\}$, $\{\exists, \wedge, \forall\}$, $\{\exists, \wedge, \vee\}$ and $\{\exists, \wedge, \forall, \vee\}$

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

(a) 4 "non-Boolean-closed" classes: $\{\exists, \wedge\}$, $\{\exists, \wedge, \forall\}$, $\{\exists, \wedge, \vee\}$ and $\{\exists, \wedge, \forall, \vee\}$

(b) 4 "Boolean-closed" classes: $\{\exists, \wedge, \neg\}$, $\{\exists, \wedge, \neg, \forall\}$, $\{\exists, \wedge, \neg, \forall, \vee\}$ and $\{\neg, \forall\}$

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

(a) 4 "non-Boolean-closed" classes: $\{\exists, \wedge\}$, $\{\exists, \wedge, \forall\}$, $\{\exists, \wedge, \vee\}$ and $\{\exists, \wedge, \forall, \vee\}$

(b) 4 "Boolean-closed" classes: $\{\exists, \wedge, \neg\}$, $\{\exists, \wedge, \neg, \forall\}$, $\{\exists, \wedge, \neg, \forall, \vee\}$ and $\{\neg, \forall\}$

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

(a) 4 "non-Boolean-closed" classes: $\{\exists, \wedge\}$, $\{\exists, \wedge, \forall\}$, $\{\exists, \wedge, \vee\}$ and $\{\exists, \wedge, \forall, \vee\}$

(b) 4 "Boolean-closed" classes: $\{\exists, \wedge, \neg\}$, $\{\exists, \wedge, \neg, \forall\}$, $\{\exists, \wedge, \neg, \forall, \vee\}$ and $\{\neg, \forall\}$

Classifying Sentences

1. Annotate the corpora/mine the patterns
2. Consider sentences expressing operators from classes $c \subseteq \{\wedge, \exists, \forall, \vee, \neg\}$ (and only such)
3. Observe the frequency of

(a) 4 "non-Boolean-closed" classes: $\{\exists, \wedge\}$, $\{\exists, \wedge, \forall\}$, $\{\exists, \wedge, \vee\}$ and $\{\exists, \wedge, \forall, \vee\}$

(b) 4 "Boolean-closed" classes: $\{\exists, \wedge, \neg\}$, $\{\exists, \wedge, \neg, \forall\}$, $\{\exists, \wedge, \neg, \forall, \vee\}$ and $\{\neg, \forall\}$

$\Rightarrow$ Each class "approximates" a fragment of English

$\Rightarrow$ Study relationships between class frequency $fr(c)$ and class rank or expressivity $rk(c)$

Corpora

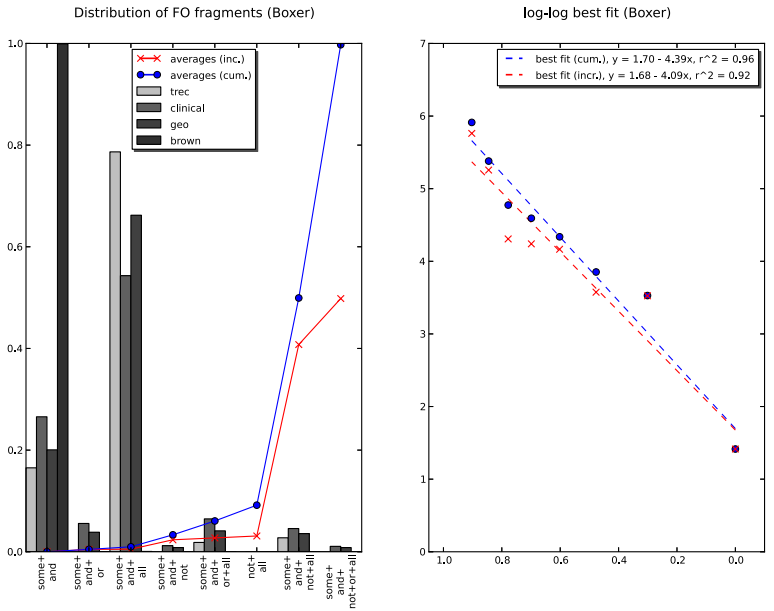
- We considered
  1. a subset (A: press articles) of the Brown corpus (http://nltk.googlecode.com/svn/trunk/nltk_data/index.xml)
  2. a subset (Geoquery880) of the Geoquery corpus (http://www.cs.utexas.edu/users/ml/nldata/geoquery.html)
  3. a corpus of clinical questions (http://clinques.nlm.nih.gov)
  4. a sample from the TREC 2008 corpus (http://trec.nist.gov)

- Of features

Corpus	Size	Domain	Type
Brown	19,741 sent.	Open (news)	Declarative
Geoquery	364 ques.	Geographical	Interrogative
Clinical ques.	12,189 ques.	Clinical	Interrogative
TREC 2008	436 ques.	Open	Interrogative

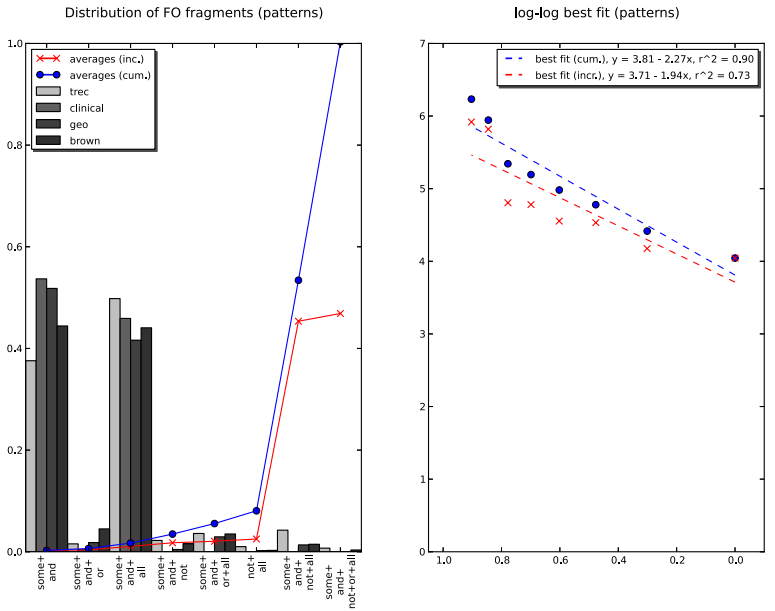
Power Law Fitting [Bar09]

Boxer



	(power law)	(R^2)
cum:	$fr(c) = \frac{1.7}{rk(c)^{4.39}}$	0.96
means:	$fr(c) = \frac{1.68}{rk(c)^{4.09}}$	0.92

Patterns



	(power law)	(R^2)
cum:	$fr(c) = \frac{3.81}{rk(c)^{2.27}}$	0.90
means:	$fr(c) = \frac{3.71}{rk(c)^{1.94}}$	0.72

Validation (Entropy, χ^2, Skewness)

- Following [Gri10], we conducted the following tests

Validation (Entropy, χ^2, Skewness)

- Following [Gri10], we conducted the following tests
  1. Computed class entropy and relative entropy

Validation (Entropy, χ^2, Skewness)

- Following [Gri10], we conducted the following tests
  1. Computed class entropy and relative entropy
  2. Run a χ^2 test and measured skewness

Validation (Entropy, χ^2, Skewness)

- Following [Gri10], we conducted the following tests
  1. Computed class entropy and relative entropy
  2. Run a χ^2 test and measured skewness

	Ent.	Rel. Ent.	Skewness	χ^2	p-value	df.
Boxer	1.53	0.51	1.93	293731473.0	0.0	7
Patterns	1.50	0.50	1.21	906727332.0	0.0	7

Validation (Entropy, χ^2, Skewness)

- Following [Gri10], we conducted the following tests
  1. Computed class entropy and relative entropy
  2. Run a χ^2 test and measured skewness

	Ent.	Rel. Ent.	Skewness	χ^2	p-value	df.
Boxer	1.53	0.51	1.93	293731473.0	0.0	7
Patterns	1.50	0.50	1.21	906727332.0	0.0	7

$\Rightarrow$ Data skewed towards simple classes

$\Rightarrow$ But... power law model inconclusive

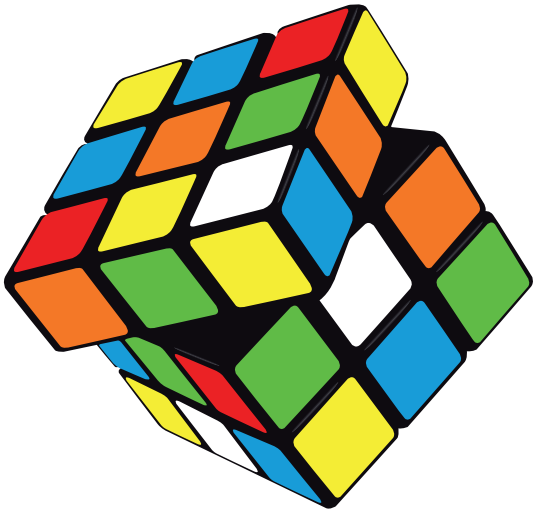
$\Rightarrow$ We don't know Boxer's accuracy either

Conclusions and Further Work

1. We have experimented with a methodology based on the Boxer semantic parser
2. The distribution obtained may seem to indicate that "non-Boolean-closed" (tractable) fragments occur more often than "Boolean-closed" (intractable) fragments
3. The results obtained thus far remain inconclusive

Conclusions and Further Work

1. We have experimented with a methodology based on the Boxer semantic parser
2. The distribution obtained may seem to indicate that "non-Boolean-closed" (tractable) fragments occur more often than "Boolean-closed" (intractable) fragments
3. The results obtained thus far remain inconclusive
4. To counter these shortcomings we would like in the future to
   - run the experiment with other deep semantic parsers (e.g., minimal recursion semantics [Cop07]),
   - consider larger corpora, in particular declarative corpora
   - consider more involved statistical tests



Thank you :-)

References I


Marco Baroni.
Distributions in text.
In Mouton de Gruyter, editor, *Corpus linguistics: An International Handbook*, volume 2, pages 803–821. 2009.


Raffaella Bernardi, Francesca Bonin, Domenico Carbotta, Diego Calvanese, and Camilo Thorne.
English querying over ontologies: E-QuOnto.
In *Proceedings of the 10th Congress of the Italian Association for Artificial Intelligence (AI*IA 2007)*, 2007.


Johan Bos.
Wide-coverage semantic analysis with Boxer.
In *Proceedings of the 2008 Conference on Semantics in Text Processing (STEP 2008)*, 2008.


Ann Copestake.
Semantic composition with (robust) minimal recursion semantics.
In *Proceedings of the ACL-07 workshop on Deep Linguistic Processing*, 2007.

References II


Norbert E. Fuchs and Kaarel Kaljurand.
Mapping Attempto Controlled English to OWL-DL.
In *Demos and Posters of the 3rd European Semantic Web Conference (ESWC 2006)*, 2006.


Stefan Th. Gries.
Useful statistics for corpus linguistics.
In Aquilino Sánchez and Moisés Almela, editors, *A mosaic of corpus linguistics: selected approaches*, pages 269–291. Peter Lang, 2010.


Richard Montague.
Universal grammar.
Theoria, 36(3):373–398, 1970.


Lawrence S. Moss.
Natural logic and semantics.
In *Proceedings of the 2009 Amsterdam Colloquium (AC 2009)*, 2010.


Reinhard Muskens.
An analytic tableau system for natural logic.
In *Proceedings of the 2009 Amsterdam Colloquium (AC 2009)*, 2010.

References III


Ian Pratt-Hartmann and Allan Third.
More fragments of language.
Notre Dame Journal of Formal Logic, 47(2):151–177, 2006.


Jakub Szymanik.
Quantifiers in Time and Space.
Institute for Logic, Language and Computation, 2009.


Camilo Thorne.
Query Answering over Ontologies Using Controlled Natural Languages.
PhD thesis, Faculty of Computer Science, 2010.

Semantic Complexity [PHT06, Tho10]

	Combined	**Data**
COP	in LSPACE	in LSPACE
COP+TV	in PTIME	in LSPACE
COP+DTV	in PTIME	in LSPACE
COP+TV+DTV	in PTIME	in LSPACE
COP+Rel	PTIME-complete	in LSPACE
COP+Rel+TV	PTIME-complete	in PTIME
COP+Rel+DTV	PTIME-complete	in PTIME
COP+Rel+DTV+TV	PTIME-complete	in PTIME

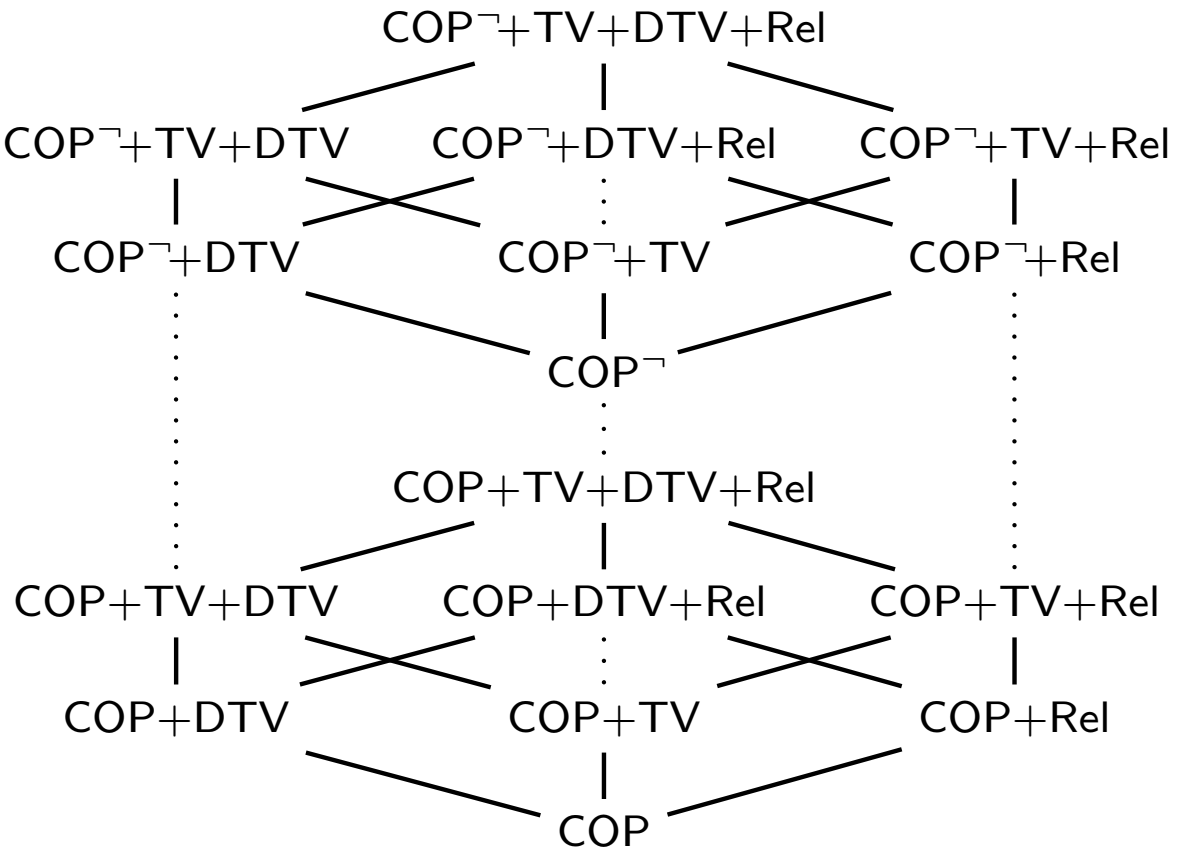
	Combined	**Data**
$\text{COP}^{\neg}$	in LSPACE	in LSPACE
$\text{COP}^{\neg}$+TV	NLSPACE-complete	in LSPACE
$\text{COP}^{\neg}$+DTV	in PTIME	in LSPACE
$\text{COP}^{\neg}$+TV+DTV	in PTIME	in LSPACE
$\text{COP}^{\neg}$+Rel	NPTIME-complete	in LSPACE
$\text{COP}^{\neg}$+Rel+TV	EXP-complete	NPTIME-complete
$\text{COP}^{\neg}$+Rel+DTV	NEXP-complete	NPTIME-complete
$\text{COP}^{\neg}$+Rel+DTV+TV	NEXP-complete	NPTIME-complete

Semantic Complexity [PHT06, Tho10]

	Combined	**Data**
COP	in LSPACE	in LSPACE
COP+TV	in PTIME	in LSPACE
COP+DTV	in PTIME	in LSPACE
COP+TV+DTV	in PTIME	in LSPACE
COP+Rel	PTIME-complete	in LSPACE
COP+Rel+TV	PTIME-complete	in PTIME
COP+Rel+DTV	PTIME-complete	in PTIME
COP+Rel+DTV+TV	PTIME-complete	in PTIME

	Combined	**Data**
$\text{COP}^{\neg}$	in LSPACE	in LSPACE
$\text{COP}^{\neg}$+TV	NLSPACE-complete	in LSPACE
$\text{COP}^{\neg}$+DTV	in PTIME	in LSPACE
$\text{COP}^{\neg}$+TV+DTV	in PTIME	in LSPACE
$\text{COP}^{\neg}$+Rel	NPTIME-complete	in LSPACE
$\text{COP}^{\neg}$+Rel+TV	EXP-complete	NPTIME-complete
$\text{COP}^{\neg}$+Rel+DTV	NEXP-complete	NPTIME-complete
$\text{COP}^{\neg}$+Rel+DTV+TV	NEXP-complete	NPTIME-complete

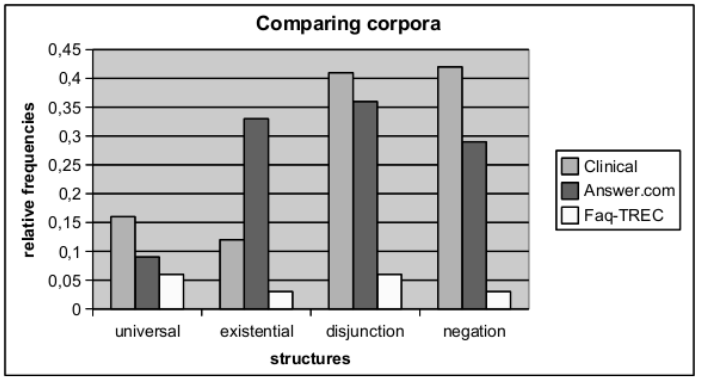
The Fragments of English (Expressive Power)



Linguistic Patters in Questions [BBC+07] (Related)

The analysis presented can be compared to the more linguistics-based methodology followed in [BBC+07]

The frequency distribution of classes of logical word patterns such as “all”, “both”, “each”, “every”, “everybody”, “everyone”, “any”, “none”, “nothing” was analyzed

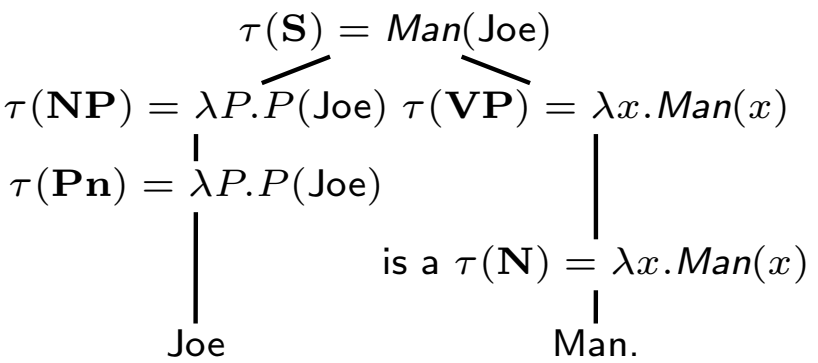


Only interrogative corpora were analyzed and co-occurence disregarded

The Fragment COP($\neg$) (Example)

Phrase Structure Rules	
$\mathbf{S} \rightarrow \mathbf{NP\ VP}$	$\tau(\mathbf{S}) = \tau(\mathbf{NP})(\tau(\mathbf{VP}))$
$\mathbf{VP} \rightarrow$ is a $\mathbf{N}$	$\tau(\mathbf{VP}) = \tau(\mathbf{N})$
$\mathbf{VP} \rightarrow$ is $\mathbf{Adj}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Adj})$
$\mathbf{VP} \rightarrow \mathbf{IV}$	$\tau(\mathbf{VP}) = \tau(\mathbf{IV})$
$\mathbf{NP} \rightarrow \mathbf{Pn}$	$\tau(\mathbf{NP}) = \tau(\mathbf{Pn})$
$\mathbf{NP} \rightarrow \mathbf{Det\ N}$	$\tau(\mathbf{NP}) = \tau(\mathbf{Det})(\tau(\mathbf{N}))$
$(\mathbf{VP} \rightarrow$ is $\mathbf{Ng}$ a $\mathbf{N}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Ng})(\tau(\mathbf{N})))$
$(\mathbf{VP} \rightarrow$ does $\mathbf{Ng\ IV}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Ng})(\tau(\mathbf{IV})))$

Function Lexicon	
$\mathbf{Det} \rightarrow$ every	$\tau(\mathbf{Det}) = \lambda P.Q.\forall x(P(x) \rightarrow Q(x))$
$\mathbf{Det} \rightarrow$ some	$\tau(\mathbf{Det}) = \lambda P.Q.\exists x(P(x) \wedge Q(x))$
$(\mathbf{Ng} \rightarrow$ not	$\tau(\mathbf{Ng}) = \lambda P.\lambda x.\neg P(x))$



Every man left $\rightsquigarrow$
$\forall x(\mathit{Man}(x) \rightarrow \mathit{Leave}(x))$

$\vdots$

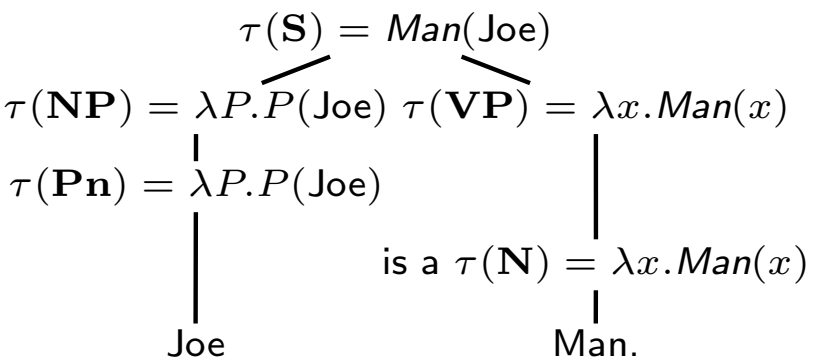
Some man is (not) a philosopher $\rightsquigarrow$
$\exists x(\mathit{Man}(x) \wedge (\neg)\mathit{Philosopher}(x))$

$\vdots$

The Fragment COP($\neg$) (Example)

Phrase Structure Rules	
$\mathbf{S} \rightarrow \mathbf{NP\ VP}$	$\tau(\mathbf{S}) = \tau(\mathbf{NP})(\tau(\mathbf{VP}))$
$\mathbf{VP} \rightarrow$ is a $\mathbf{N}$	$\tau(\mathbf{VP}) = \tau(\mathbf{N})$
$\mathbf{VP} \rightarrow$ is $\mathbf{Adj}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Adj})$
$\mathbf{VP} \rightarrow \mathbf{IV}$	$\tau(\mathbf{VP}) = \tau(\mathbf{IV})$
$\mathbf{NP} \rightarrow \mathbf{Pn}$	$\tau(\mathbf{NP}) = \tau(\mathbf{Pn})$
$\mathbf{NP} \rightarrow \mathbf{Det\ N}$	$\tau(\mathbf{NP}) = \tau(\mathbf{Det})(\tau(\mathbf{N}))$
$(\mathbf{VP} \rightarrow$ is $\mathbf{Ng}$ a $\mathbf{N}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Ng})(\tau(\mathbf{N}))$
$(\mathbf{VP} \rightarrow$ does $\mathbf{Ng\ IV}$	$\tau(\mathbf{VP}) = \tau(\mathbf{Ng})(\tau(\mathbf{IV}))$

Function Lexicon	
$\mathbf{Det} \rightarrow$ every	$\tau(\mathbf{Det}) = \lambda P.Q.\forall x(P(x) \rightarrow Q(x))$
$\mathbf{Det} \rightarrow$ some	$\tau(\mathbf{Det}) = \lambda P.Q.\exists x(P(x) \wedge Q(x))$
$(\mathbf{Ng} \rightarrow$ not	$\tau(\mathbf{Ng}) = \lambda P.\lambda x.\neg P(x))$



Every man left $\rightsquigarrow$
$\forall x(Man(x) \rightarrow Leave(x))$

$\vdots$

Some man is (not) a philosopher $\rightsquigarrow$
$\exists x(Man(x) \wedge (\neg) Philosopher(x))$

$\vdots$

Entropy and Relative Entropy (Reminder)

The entropy $H(X)$ of a (random) variable $X = x_1, \ldots, x_n$ is

$$H(X) = -\sum_i p(x_i) \cdot \log_2(p(x_i))$$

where $p(\cdot)$ is the p.m.f. of X; entropy measures how random the distribution of X is

The relative entropy $H_{rel}(X)$ of X, defined as

$$H_{rel}(X) = \frac{H(X)}{\log_2(n)}$$

and transforms $H(X)$ into a metric, since

$$\log_2(n) = \log_2(\max_i p(x_i))$$

Power Laws and Log-Log Regressions (Reminder)

We can transform power law models into linear models via logarithmic scaling

$$
\begin{aligned}
y &= \frac{b}{x^m} \\
&\Leftrightarrow \\
\log_{10}(y) &= \log_{10}(\tfrac{b}{x^m}) \\
&= \log_{10}(b) - \log_{10}(x^m) \\
&= \log_{10}(b) - m \cdot \log_{10}(x)
\end{aligned}
$$

If the R^2 coefficient is sufficiently high, we can say (with some caveats!!!) that the data (sample) is power law distributed

Linear Regression (Reminder)

A linear regression model has the form

$$Y = \theta X$$

with parameters $\theta = (m, b)^T$ (a gradient and an intercept)

It can be used to predict the value(s) of a dependent variable Y (e.g., frequency) vis-á-vis the value(s) of a predictor X (e.g., class complexity rank)

The least squares method computes the linear model whose parameters θ^* minimize square error

$$\theta^* = \arg\min_{\theta} \sum_i (y_i - \theta(x_i))^2$$

The R^2 coefficient provides a measure of confidence in the inferred model $Y = \theta^* X$ and is defined in terms of square error variance and dependent variable variance, i.e.,

$$R^2 = \frac{Var(\theta^* X)}{Var(Y)}$$