4. Aggregate Quantifier Distribution

Camilo Thorne[1]

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

ESSLLI 2013, Aug 2-6, Düsseldorf

Outline

http://www.inf.unibz.it/~cathorne/agg_esslli

Motivation

- Words and structures in English occur following some general laws
- A distribution describes how often they occur/probable they are
- E. Zipf showed that in many cases such distributions correspond to power law
- We want to know if this applies to quantifiers too

The Fragments of English (FOEs) [PHT06]

COP	Copula, common and proper nouns, negation, universal, existential quantifiers
COP+Rel	COP plus relative pronouns
COP+TV	COP plus transitive verbs
COP+TV+DTV	COP+TV plus ditransitive verbs
COP+Rel+TV	COP+Rel plus transitive verbs
COP+Rel+TV+DTV	COP+Rel+TV plus ditransitive verbs
COP+Rel+TV+RA	COP+Rel+TV plus anaphoric pronouns (e.g., he, him, it, herself) of bounded scope
COP+Rel+TV+GA	COP+Rel+TV plus unbounded anaphoric pronouns
COP+Rel+TV+DTV+RA	COP+Rel+TV+DTV plus bounded anaphoric pronouns

Engendered Logics and Complexity

- Modulo $\tau(\cdot)$ controlled fragments express (translate exactly into) a fragment of FO

- COP expresses the FO fragment containing the following finite set of sentence forms:

$\mathit{Woman}(\mathrm{Mary})$	Mary is a woman.
$\neg \mathit{Man}(\mathrm{Mary})$	Mary is not a man.
$\forall x(\mathit{Man}(x) \Rightarrow \mathit{Person}(x))$	Every man is a person.
$\forall x(\mathit{Woman}(x) \Rightarrow \neg \mathit{Man}(x))$	No woman is a man.
$\forall x(\mathit{Person}(x) \Rightarrow \mathit{Human}(x))$	Every person is a human.
$\forall x(\mathit{Person}(x) \Rightarrow \mathit{Human}(x))$	Every person is a human.
$\exists x(\mathit{Person}(x) \wedge \mathit{Woman}(x))$	Some person is a woman
$\exists x(\mathit{Person}(x) \wedge \neg \mathit{Woman}(x))$	Some person is not a woman.

Engendered Logics and Complexity

- Modulo $\tau(\cdot)$ controlled fragments express (translate exactly into) a fragment of FO

- COP expresses the FO fragment containing the following finite set of sentence forms:

$\mathit{Woman}(\mathrm{Mary})$	Mary is a woman.
$\neg \mathit{Man}(\mathrm{Mary})$	Mary is not a man.
$\forall x(\mathit{Man}(x) \Rightarrow \mathit{Person}(x))$	Every man is a person.
$\forall x(\mathit{Woman}(x) \Rightarrow \neg \mathit{Man}(x))$	No woman is a man.
$\forall x(\mathit{Person}(x) \Rightarrow \mathit{Human}(x))$	Every person is a human.
$\forall x(\mathit{Person}(x) \Rightarrow \mathit{Human}(x))$	Every person is a human.
$\exists x(\mathit{Person}(x) \wedge \mathit{Woman}(x))$	Some person is a woman
$\exists x(\mathit{Person}(x) \wedge \neg \mathit{Woman}(x))$	Some person is not a woman.

- By studying FO we can understand the complexity of semantic processing $\Rightarrow$ semantic complexity

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	$\forall x(\mathit{Man}(x) \wedge \neg \mathit{Cheat}(x)$ $\rightarrow \mathit{Trusted}(x))$
$\vdots$	$\vdots$	$\vdots$

Complexity of the FOEs [PHT06]

COP	in NL
COP+TV	NL-complete
COP+TV+DTV	in P
COP+Rel	NP-complete
COP+TV+Rel	EXP-complete
COP+TV+Rel+RA	EXP-complete
COP+TV+DTV+Rel	NEXP-complete
COP+TV+Rel+GA	undecidable

Different function words yield different complexity!

Complexity of the FOEs [PHT06]

COP	in NL
COP+TV	NL-complete
COP+TV+DTV	in P
COP+Rel	NP-complete
COP+TV+Rel	EXP-complete
COP+TV+Rel+RA	EXP-complete
COP+TV+DTV+Rel	NEXP-complete
COP+TV+Rel+GA	undecidable

Not "Boolean-closed" $\Rightarrow$ tractable!

Complexity of the FOEs [PHT06]

COP	in NL
COP+TV	NL-complete
COP+TV+DTV	in P
COP+Rel	NP-complete
COP+TV+Rel	EXP-complete
COP+TV+Rel+RA	EXP-complete
COP+TV+DTV+Rel	NEXP-complete
COP+TV+Rel+GA	undecidable

"Boolean-closed" $\Rightarrow$ intractable!

Complexity of the FOEs [PHT06]

COP	in NL
COP+TV	NL-complete
COP+TV+DTV	in P
COP+Rel	NP-complete
COP+TV+Rel	EXP-complete
COP+TV+Rel+RA	EXP-complete
COP+TV+DTV+Rel	NEXP-complete
COP+TV+Rel+GA	undecidable

"Boolean-closed" + restricted anaphora $\Rightarrow$ decidable!

Complexity of the FOEs [PHT06]

COP	in NL
COP+TV	NL-complete
COP+TV+DTV	in P
COP+Rel	NP-complete
COP+TV+Rel	EXP-complete
COP+TV+Rel+RA	EXP-complete
COP+TV+DTV+Rel	NEXP-complete
COP+TV+Rel+GA	undecidable

"Boolean-closed" + full anaphora $\Rightarrow$ undecidable!

Power Law Distributions [Bar09]

Definition (Power law)

We say that a random variable X of outcomes $x_1, \ldots, x_k$ follows a power law or Zipf distribution if approx. 20% of its outcomes concentrate approx. 80% of its probability pass. This relation is described by the equation:

$$P(x) \sim \frac{b}{\mathit{rank}(x)^m}$$

We want to know if quantifier distribution $P(Q)$ is power-law correlated to quantifier expressiveness/complexity:

$$P(Q) \sim \frac{b}{\mathit{comp}(Q)^m}$$

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

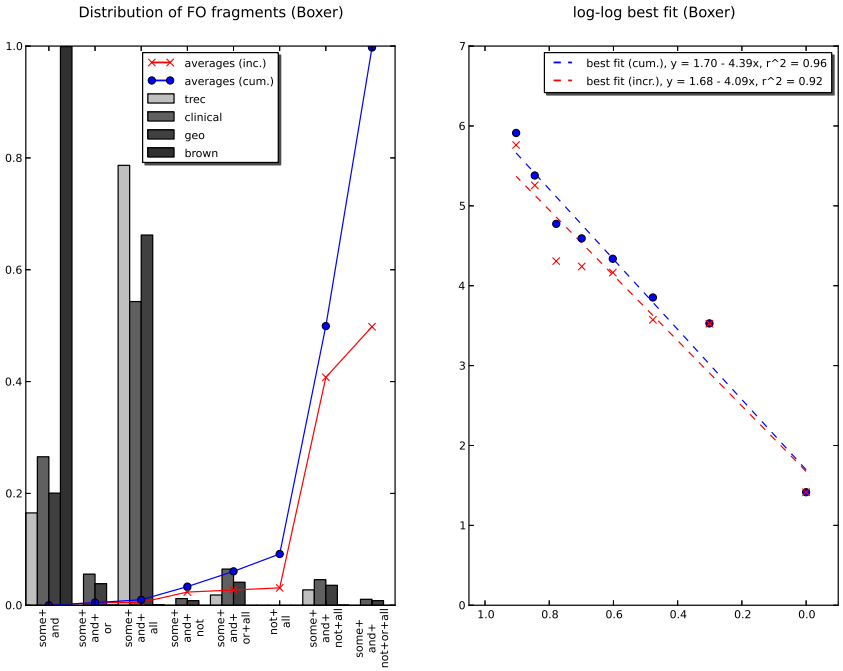
$$\Theta^* = \arg\min_{\Theta} J(\Theta) = \arg\min_{\Theta} \sum_i (y_i - \Theta(x_i))^2$$

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.,

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

Power Law Fitting [Tho12]

Boxer

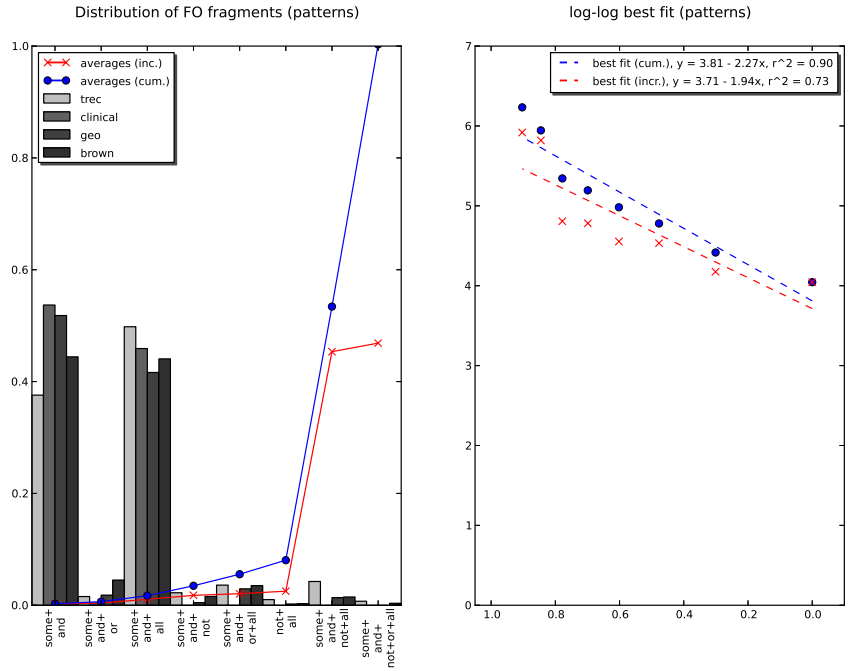


(power law) (R^2)

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

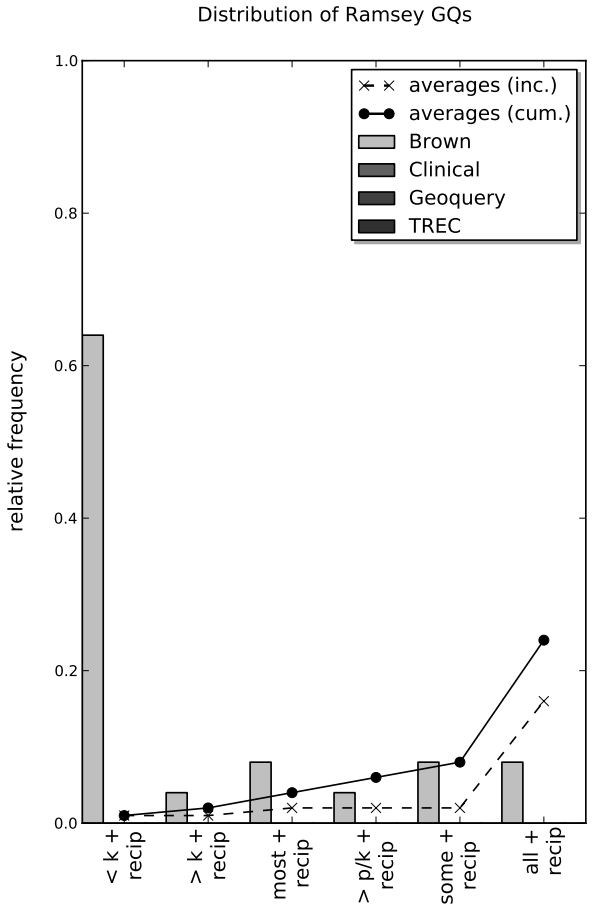
Corpora

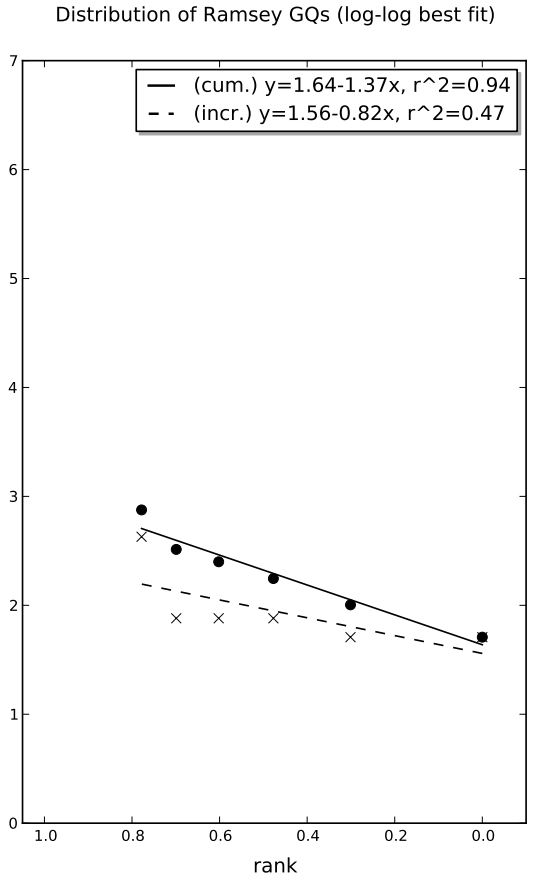
Corpus	Size	Domain	Type
Brown	19,741 sentences	Open (news)	Declarative
Geoquery	364 questions	Geographical	Interrogative
Clinical ques.	12,189 questions	Clinical	Interrogative
TREC 2008	436 questions	Open	Interrogative

Remark

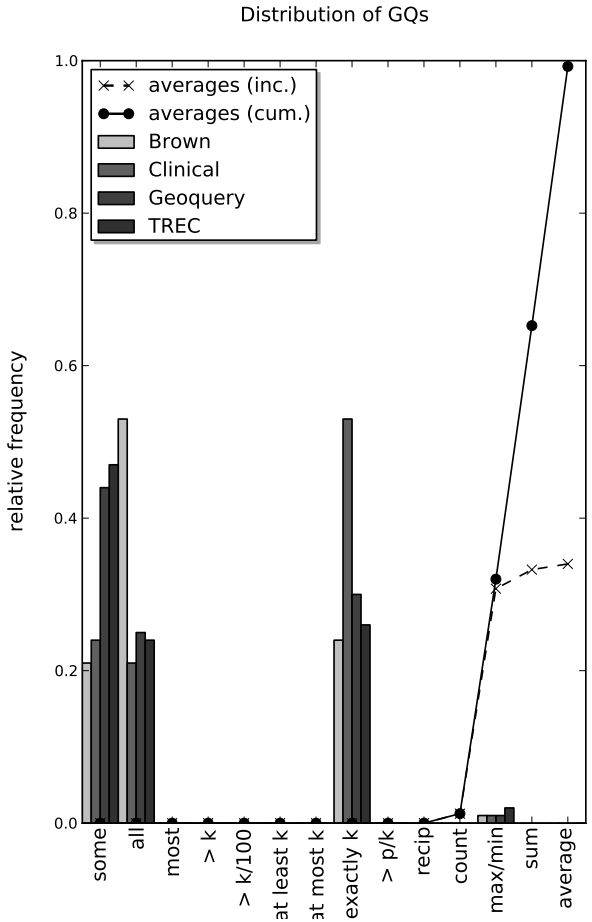
Corpora of different types and domains and approx. 1,000,000 words (cumulatively)

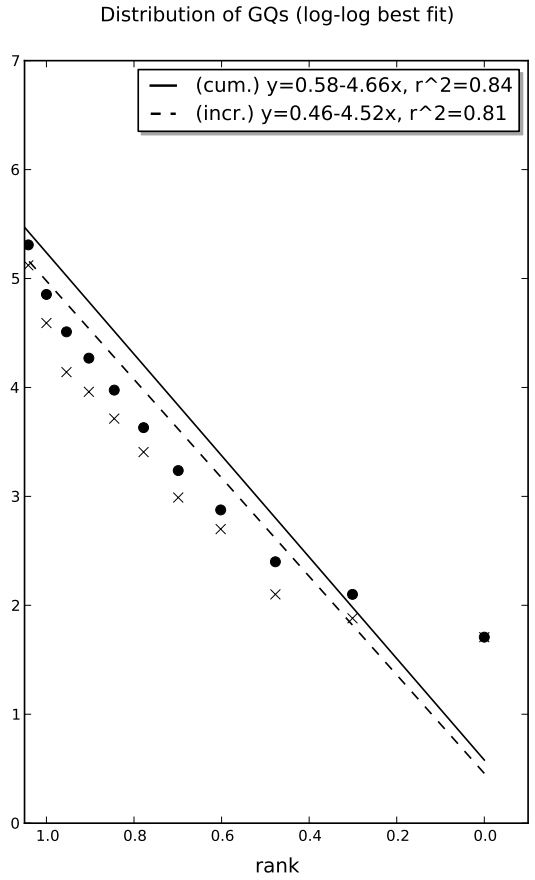
Ramsey Quantifier Distribution [TBC13]





Quantifier Distribution [TBC13]





Ramsey and non-Ramsey (raw) [TBC13]

Corpus	$> k+$ recip	$> p/k+$ recip	most+ recip	some+ recip	all+ recip	$< k+$ recip
Brown	1	1	2	2	2	16
TREC	0	0	0	0	0	0
Geo	0	0	0	0	0	0
Clin. qs.	0	0	0	0	0	0
total	1	1	2	2	2	16

Corpus	$\geq k$	$\leq k$	most	$> k$	$> p/k$	recip.	$> k\%$	sum	cnt	avg	max/min	all	k	some
Brown	192	4	1532	540	38	101	2	1	354	17	4368	202587	90811	81693
TREC	0	0	0	0	0	0	0	0	0	0	13	192	490	222
Geo	2	0	0	0	0	0	0	0	1	0	18	380	447	660
Clin. qs.	12	0	28	12	0	0	0	0	9	2	889	10712	11629	20780
total	206	4	1560	552	38	101	2	1	364	19	5288	213871	103377	103355

Test Statistics

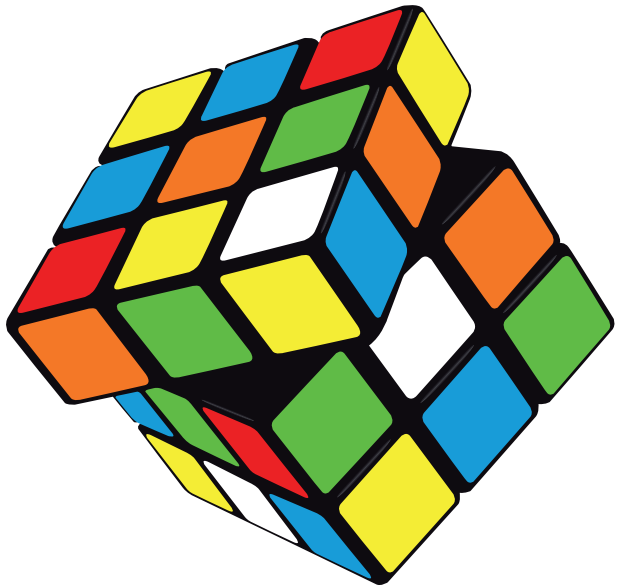
Test	Recip. GQs	GQs
Skewness	1.76	1.98
χ^2 value	530.81	183815415173.11
p value, df.	1.78, 5	0.0, 13
P. law $fr(Q)$	$36.00/rk(Q)^{-0.82}$	$2.88/rk(Q)^{-4.52}$
R^2 coeff.	0.86	0.81

Remark

The power law refers to mean relative frequency

Summary

1. We have studied the distribution of generalized quantifiers in corpora
2. We have considered if such distribution can be modeled by a power law
3. We have compared results for quantifiers with results for other
4. It may seem that distributions are skewed towards low complexity quantifiers
5. More in general, it may seem distributions are skewed towards low complexity constructs
6. This is consistent with cognitive experiments done by cognitive scientists [TBC13]



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.


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


Camilo Thorne, Raffaella Bernardi, and Diego Calvanese.
Designing efficient controlled languages for ontologies.
In Harry Bunt, editor, *Computing Meaning 6*. Springer, 2013.


Camilo Thorne.
Studying the distribution of fragments of english using deep semantic annotation.
In *Proceedings of the ISA-9 Workshop*, 2012.