Corpus pragmatics
- a tutorial

10/15/2014

Judith Degen

based on:
Degen (under review). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods.
General steps

1. formulate a research question

2. corpus search

3. additional annotation

4. data analysis and visualization
From question to answer

1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?

2. If there is variation among implicatures, is it random or systematic?
1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?
1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?
1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?
2. If there is variation among implicatures, is it random or systematic?

**Partitive**

- Simple vs. Partitive

**Subjecthood**

- Other vs. Subject

**Linguistic mention**

- New, inferable, mentioned
Methodology

- Corpus search
  - extract instances of “some”
  - extract information about syntactic/semantic/pragmatic context
- Web-based experimentation
  - collect implicature strength judgments
- Visualization and statistical data analysis
Why corpora?

Doesn’t intuition suffice? Scalar implicatures from some to not all easily survive context shifts. Homogeneous.

(1) Ann: Was the exam easy?
   Tom: Some of the students failed.
   $\leadsto$ Some, but not all, of the students failed.

(2) Ann: How is the teacher doing?
   Tom: Some of the students failed.
   $\leadsto$ Some, but not all, of the students failed.
Problems with intuitions

see also Gibson et al. (2011)

• just one participant (researcher)

• just a handful of items (hand-selected by researcher)

• consequences:
  • bias in judgments / selection of items calls into question generalizability of resulting theories
  • unclear what contexts some actually occurs in
Advantages of corpora

see also de Marneffe & Potts (2014)

• naturalistic data as opposed to (only) made-up examples

• avoids problem of exposing participants to potentially unnatural distributions (as is often the case in balanced psycholinguistic studies)

• complements intuition-based theorizing and controlled psycholinguistic experimentation
Selecting a corpus

• spoken vs. written language

• genre

• size

• available annotation
  • POS-tagged, syntactically parsed
  • coreference annotation
  • information status of NPs
  • animacy
  • prosodic / phonetic / phonological annotation
Switchboard Corpus


• spoken American English

• telephone dialogs between strangers about pre-defined topics

• ~ 800,000 tokens

• POS-tagged, syntactically parsed; information status annotation for ~ 23% of NPs Nissim et al. (2004)
Extract data from corpus

• use tgrep2 Rohde (2005) and the TGrep2 Database Tools (TDT) Degen & Jaeger (2011) to construct a database of 1749 “some” utterances

• example: add information about partitive and grammatical function

DEMO
Exclusion

• Cases where NP head is sg count noun (359):
  (1) She stuck my name on some list.
      * She stuck my name on some, but not all, list.
  (2) John kicked some cat off the street.
      ? John kicked some, but not all, cat off the street.

• Cases where entire NP consists of some (26):
  (3) Some say that coffee is healthy.

• Leaves 1363 cases
Collecting implicature strength ratings
Collecting implicature strength ratings

• Amazon’s Mechanical Turk crowd-sourcing service

• for each item, collected similarity rating on 7-point Likert scale

• blocks of 20 items, 10 ratings per item (243 participants)

• 2 practice items

https://www.hlp.rochester.edu/mturk/jdegen/7_qpsome/output/qp.html?assignmentId=foo&list=3
Analysis

1. Do scalar implicatures from some to not all constitute a homogeneous class of inferences?
Analysis

1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?
Analysis

1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?

![Histogram of Mean Rating by Item]
Analysis

2. If there is variation among implicatures, is it random or systematic?
2. If there is variation among implicatures, is it random or systematic?
Analysis

2. If there is variation among implicatures, is it random or systematic?
2. If there is variation among implicatures, is it random or systematic?
2. If there is variation among implicatures, is it random or systematic?
Analysis

Partitive                    Subjecthood               Linguistic mention

Is it all the same effect? Eg, discourse accessibility? Or are the effects independent?
Mixed effects linear regression

\[ \text{Rating}_{ij} = \beta_0 + \beta_1 \text{Partitive}_{ij} + \beta_2 \text{GrammaticalFunction}_{ij} + \beta_3 \text{InfoStatus}_{ij} + b_j + \epsilon_{ij} \]

by-participant differences \( \sim \mathcal{N}(0, \sigma_b) \)

noise \( \sim \mathcal{N}(0, \sigma_\epsilon) \)

\[ m = \text{lmer}(\text{Rating} \sim \text{cPartitive} + \text{cGrammaticalFunction} + \text{cInfoStatus} + (1|\text{workerid}), \text{data=centered}) \]

\[ \text{summary}(m) \]
Linear mixed model fit by REML
Formula: Rating ~ cPartitive + cGrammaticalFunction +
cInfoStatus + (1 | workerid)
   Data: centered
AIC   BIC logLik deviance REMLdev
56405 56450 -28197    56375 56393
Random effects:
 Groups   Name        Variance Std.Dev.   
     workerid (Intercept) 0.47074 0.68611
             Residual 3.55331 1.88502
Number of obs: 13630, groups: workerid, 243

Fixed effects:
                 Estimate Std. Error t value  
      (Intercept)     3.96828    0.04989   79.55
     cPartitive      1.16780    0.03861   30.25
cGrammaticalFunction  0.85315    0.04396   19.41
     cInfoStatus     0.41245    0.03564   11.57
Model evaluation

Full model

Subject variability

\[
R^2_{\text{marginal}} = .16
\]
\[
R^2_{\text{conditional}} = .27
\]

\[
R^2_{\text{marginal}} = 0
\]
\[
R^2_{\text{conditional}} = .09
\]
Conclusions

1. Do scalar implicatures from *some* to *not all* constitute a homogeneous class of inferences?

No.

2. If there is variation among implicatures, is it random or systematic?

The variation is systematic: implicature strength is dependent on various contextual features. But there is quite some residual variation to be explained!
Tools used

• extracting data from corpus
tgrep2 / TDT

• setting up mturk experiment
  javascript / HTML / mturk command-line tools

• data analysis & visualization
  R (especially lmer and ggplot)

• general pre- and post-processing
  python / bash
Resources

- TGrep2 User Manual
  http://tedlab.mit.edu/~dr/Tgrep2/tgrep2.pdf

- TGrep2 Tutorial
  http://www.stanford.edu/dept/linguistics/corpora/cas-tut-tgrep.html

- TGrep2 Database Tools (TDT) User Manual

- Sample experiment (change list parameter for additional items)
  https://www.hlp.rochester.edu/mturk/jdegen/7_qpsome/output/qp.html?assignmentId=foo&list=1