Constructing quantitative grammatical arguments

Aaron Ecay

University of Pennsylvania

Oct. 15, 2014
Goals

▶ Goals of this talk:
  ▶ Review foundational literature on quantitative arguments in historical syntax
  ▶ Motivate diachronic connection between use and grammar
  ▶ Provide students with tools for conducting such analyses on their own
    ▶ Not a methodological talk, however
  ▶ Interactivity, “interruptions,” etc. encouraged – if you have a question, ask!
Outline

Introduction
  Introduction

The CRH (Kroch 1989)
  Background
  Case study 1: French V2
  Case study 2: *do*-support

A critical evaluation
  ME negation
What is the CRH?

► **Constant Rate Hypothesis**: changes spread at the same rate in all contexts.

► “On the basis of [this hypothesis] substantial progress can be made in understanding the relationship between the structural patterns uncovered by grammatical analysis and the frequency patterns revealed by sociolinguistic methods.”

► Kroch (1989)
Why the CRH?

- The CRH is fundamentally a parsimony argument
  - Ockham’s razor
  - “Methodological minimalism” if you use Chomsky for your philosophy of science
The CRH and parsimony

- There are two possible parameterizations of a system of parallel lines
  - 4 parameters: \( \langle \text{slope, intercept} \rangle \times 2 \)
  - 3 parameters: \( \langle \text{slope, intercept, offset} \rangle \)
- The parsimony gains increase as more lines are added to the system
- The CRH says: take the more parsimonious description
Terms and definitions

- Some terminological clarification
  - **parameter**: discrete choice that a grammar makes
  - **grammar**: how a speaker decides to structure their utterances. Borer-Chomsky Conjecture: a list (lexicon) of functional items with features (Minimalism, HPSG, ...)
  - **competition**: the ability of native speakers to learn/process/produce sentences from a variety of grammars. Cf. balanced bilinguals...
CRH and evolution

- Viewing language variation as competition allows the emergence of interesting models from population biology (Yang 2000)
- The rate of spread of an innovation is proportional to the number of speakers who have the innovation, and the number of speakers who don’t
- This is the same functional form that describes the progress of an invasive species in a closed ecosystem (e.g. an island)
  - Namely, the logistic curve
CRH methodology

- Logistic regression fits logistic curves to data (R, Varbrul, ...)
- Result of logistic regression
  - Intercept and slope for each context
  - $p$-value for the hypothesis that each slope differs from zero
    - If this $p$-value is large, you can drop that slope term
    - Ideally all $p$-values are large → you are in the simple 1 slope, n intercepts model
  - Other model comparison methods are possible (and preferable)
Preliminaries

- This discussion is based on Kroch’s treatment
- There is research underlying this, cited by Kroch and more recent
French V2

- Old French is a verb-second language
  - When an object appears pre-verbally, the subject must appear post-verbally
  - The same rule applies, optionally, to fronted PPs and adverbs
    - But V2 is still general: French (unlike modern Gmc) has a leftward-adjointed position that doesn’t trigger inversion
  - Ignore clitics (including subject clitics!)
French V2 and null subjects

- Older French had null subjects, but only postverbally
  - That is: when V2 inversion had applied
- Loss of V2 → loss of null subjects
French V2: data

Reconstructed from Kroch (1989) Figure 3; underlying data from Fontaine (1985)
The CRH in French V2 data

- The three contexts are well-modeled by logistic curves
- All three have the same slope
- The availability of null subjects and of V2 are controlled by the same grammatical parameter
French V2: going deeper

- Old French clause: \([_{LFD} (DP) |_IP [_{TOP} DP V ]_]\)
- A change in prosody leads to the modern French situation: one stress per intonation phrase (IP)
- Don’t topicalize, left-dislocate
  - Leads to apparent surface violations of V2 constraint
French LFD data

variable
- LFD (pred)
- NP-V2
- null subj
- pro-V2
French V2 and topicalization

- We can measure the new prosody by counting the number of left-dislocations (= leftward movement that leaves a clitic behind)
- It is parallel to the V2 and null subject lines
We can measure the new prosody by counting the number of left-dislocations (= leftward movement that leaves a clitic behind)

It is parallel to the V2 and null subject lines

A change in prosody causes the loss of V2, which causes the loss of null subjects (!)
French V2 and topicalization

- We can measure the new prosody by counting the number of left-dislocations (= leftward movement that leaves a clitic behind)
- It is parallel to the V2 and null subject lines
- A change in prosody causes the loss of V2, which causes the loss of null subjects (!)
- Really?
French V2 and prosody: data

Here are the data on the original scale
French V2 and prosody: a critical perspective

- Logistic regression slope = how long does this change take?
- Many historical changes take approximately the same amount of time
- $p$-value (traditional decision criterion for logistic regression) $\approx$ measure of sample size
French V2 and prosody: a critical perspective

- Logistic regression slope = how long does this change take?
- Many historical changes take approximately the same amount of time
- \( p \)-value (traditional decision criterion for logistic regression) \( \approx \) measure of sample size
- This introduces “researcher degrees of freedom” into analyses (Simmons et al. 2011)
French V2 and prosody: constructive criticism

- I’m not saying these results are doubtful
  - Because they have independent support from non-quantitative analyses
  - The change in prosody could have happened faster, later, ...
- Quantitative data on their own don’t (dis)prove anything
  - Just like non-quantitative data
- Quantitative analysis generates observational facts that grammatical theories must cope with
English *do*-support

- Use of semantically vacuous auxiliary *do* in certain morphosyntactic contexts in English
- Develops in Early Modern English (~1500–1700)
Data from Ellegård (1953)
Explication of data

- *do*-support rises in various contexts; some before others
- Something happens to the monotonic upwards trajectory in 1575
  - ignore data after this date
  - (see Warner 2005, Ecay 2014)
Grammatical explanation

- Posited link between *do*-support and verb raising
- When verbs no longer raise, *do*-support becomes necessary to support stranded affix in T (Embick and Noyer 2001)
**do-support and verb raising**

- **Measuring verb raising:**
  
  (1)  
  a. Queen Esther looked never with such an eye  
  b. Queen Esther never looked...

- By the CRH, the loss of the construction in (1a) should be parallel to the emergence of *do*-support
  
  - They are controlled by the same underlying parameter: ± V-to-T
Measuring verb raising: a bump in the road

- However, sometimes *never* is left-adjointed to T:
  
  (2) John never will find out the secret

- This word order is rare but grammatical since ME
  - Kroch (1989) finds 10–18% usage without the benefit of a parsed corpus
  - Measuring more carefully in a parsed corpus nets a lower estimate (3–6%)

- Thus, we can have an apparently ModE word order even with V-to-T:

```
TP
  DP
    | Queen Esther
  T'
    | AdvP
      | never
  T
    | V
      | looked
  T
    | t_v
      | with such an eye
VP
  PP
```
Measuring verb raising: the solution

- Thus, we should disregard 16% of the observed tokens of *never V word order*
- How?
  - Kroch binned the data, so he just multiplies each bin’s total by 0.84 (= 1 - 0.16)
  - Binning is bad practice, though
    - Bootstrap
    - Custom model (Bayesian: JAGS, STAN, ...)
The data on *do*-support and verb raising across *never* provide no evidence against the common-slope hypothesis.

Conclusion: the CRH applies here

- Strengthens the hypothesis that both these surface phenomena are controlled by a ± V-to-T parameter
Middle English negation

- In Middle English, there is a change in the exponence of Neg.
- The negator *ne*, inherited from OE, is lost.
- *not*, formerly a negative adverb, becomes the new negator.
Details of the change

During the period of the change, a large number of negative sentences have both *ne* and *not*:

(3) he ne shal nouȝt decieue him

Frisch (1997)

- Frisch examines this situation and concludes:
  - There are two grammars of negation
    - $ne = \text{Neg}^0$
    - $not = \text{Spec, NegP}$
  - Since these are not mutually incompatible, they don’t compete, but rather cooperate to generate $ne + not$ sentences
Correcting for adverbial *not*

- Some uses of *not* are adverbial, not sentence negation
- Diagnosable sometimes by position

(4) Þat Jesuss nohht ne wollde ben boren nowwhar i þe land
that J. not NEG would be born nowhere in the land
‘That Jesus did not want to be born anywhere in the land’
*Ormulum*, from Frisch (1997)

- Same pre-T position used by *never*
  - Same 0.16 correction applies
Frisch’s conclusions

- Frisch concludes that *ne* and *not* are not in competition, since their slopes differ numerically.
  - If they were in competition, one slope would be the negative of the other.

- He also concludes that the two grammars are independent, since
  \[ P(ne\ldots not) \approx P(ne) \times P(not) \]
Wallage (2008)

- Advocates a more complicated model
  - \( ne_{[-\text{neg}]} \), \( ne_{[-\text{neg}]} \), \( not_{[-\text{neg}]} \)
- Proposes several empirical refinements to Frish’s model
  - Split between subordinate and main clauses
    - Frish’s model fits main clauses better
  - Better controls on independence assumptions
    - “double counting”
Wallage’s stage one and the CRH

- Tests the CRH with respect to the data (in an odd and insufficient way)
- His conclusion: loss of $ne_{[+neg]}$ (stage one) obeys the CRH
- We can (probably) agree

<table>
<thead>
<tr>
<th>Period</th>
<th>Input</th>
<th>Main cls</th>
<th>Sub cls</th>
<th>if-cls</th>
<th>Scope of neg</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1250–1350</td>
<td>0.712</td>
<td>0.288</td>
<td>0.701</td>
<td>–</td>
<td>–</td>
<td>0.0001</td>
</tr>
<tr>
<td>1350–1420</td>
<td>0.01</td>
<td>0.250</td>
<td>0.660</td>
<td>0.936</td>
<td>0.963</td>
<td>0.0001</td>
</tr>
<tr>
<td>1420–1500</td>
<td>0.003</td>
<td>0.243</td>
<td>0.717</td>
<td>0.921</td>
<td>0.965</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Wallage’s stage two and the CRH

- Wallage finds no significant difference between contexts for the loss of $ne_{[-neg]} = ne...not = “stage two”$
- His conclusion: no CRH
- Our conclusion: CRH to the max
What happened here?

- There were two theories of the syntax of *ne*
  - Frisch: maximally simple
  - Wallage: more complex
- With a small amount of data and analysis, the simple solution looks correct
- More data and analysis make the complex theory look better (see also Ecay and Tamminga 2013)
  - You don’t get to have a more complicated theory without data to match
- Seems like science!
What’s the lesson?

- Use the best grammatical theory you have available
- Don’t be afraid to be (eventually) proven wrong
- (Share your data and methods!)


Bibliography II


