Does information density help?
A Bayesian analysis of $help + (to) \, V_{\text{INF}}$
in geographic varieties of English

Natalia Levshina
Leipzig University

DGfS Meeting, Saarbrücken, March 2017
Outline

1. help (to) Vinf:
   • Previous research
   • Information density hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
   • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
Alternation

• Mary helped John cook the dinner.
• Mary helped John to cook the dinner.

• A rare case when variation is possible between the bare and marked infinitive in PDE.
• Mair (2002) and Rohdenburg (2009): diachronic evidence that the bare infinitive has been gradually taking over, at least since the mid-19th century, with AmE leading.
When marked, when bare?

• Dixon (1991: 199): bare Vinf means a more active involvement of the Helper:
  • John helped Mary eat the pudding (he ate half).
  • John helped Mary to eat the pudding (by guiding the spoon to her mouth, since she was still an invalid).

• Not all agree with that (e.g. Huddleston & Pullum 2002: 1244).
Complexity

• Principle of cognitive complexity:
  “In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favoured in cognitively more complex environments” (Rohdenburg 1996: 151).

• The greater the distance between help and Vinf, the higher the chances of the marked infinitive.
  • I helped (them) as well as I could to wash up.
  • ?? I helped (them) as well as I could wash up. (Rohdenburg 1996: 159).

• Statistical evidence in Lohmann (2011), based on the BNC.
Horror aequi

• Avoidance of identity (repetition in near context)

  • Sorry, but how is this supposed to help answer the question?
  • ?? Sorry, but how is this supposed to help to answer the question?

• Interacts with formal distance: the greater the distance between help and Vinf, the weaker the effect (Lohmann 2011).
Presence of Helpee

• the bare infinitive is particularly dominant in the pattern help + NP + Vinf (Biber et al. 1999: 735; Lohmann 2011).

• Vegetable soup helps you lose weight.
• The mic stand helps to get the microphone placed properly for the best sound quality possible.
Inflectional form

• The form *helping* favours *to-Vinf* (Lohmann 2011).

  • I look forward to Vicky *helping* me *to buy* more clothes next season!
Register

• help + bare Vinf is particularly preferred in informal registers (Biber et al. 1999: 736–737).
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis
2. GloWbE corpus
3. Variables:
   • Traditional
   • Information-theoretic
4. Bayesian mixed-effect logistic regression models
5. Discussion
Uniform Information Density

• Information Theory and hypothesis of Uniform Information Density (Jaeger 2010):
  • Less informative/more predictable units tend to require less coding
    • bare Vinf in more predictable contexts?
  • More informative/less predictable units tend to require more coding
    • marked Vinf in less predictable contexts?
OK, but what is predictability?

• Defining the context: different levels of abstraction

Predictability given the immediate lexical context (words/ngrams), e.g. Piatandosi et al. (2011) or

Predictability given syntactic information, e.g. Jaeger (2010): predictability of complement clauses given matrix verbs helps predict the use or omission of *that*, e.g. *I know he did it* vs. *I read that he did it*.
OK, but what is predictability?

• Left of right context (in case of ngrams)?
  • E.g. Bell et al. (2009): left context is more important for phonetic reduction of function words, while right context is more important for reduction of content words.

• Direction of predictability:
  • predictability of $V_{inf}$ given the context? Or predictability of the context given $V_{inf}$?
Recall the intro talk...
Research questions

• Does information density help in general to model the variation of help of help?
• Which types of informativity/predictability are (more) important wrt. this variation?
• Do we find similar tendencies in different varieties of English?
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
   • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
GloWbE corpus (Davies 2013)
Data

• Online English from 20 countries
• Created by Google Advanced Search [Region] for highly frequent ngrams from COCA
• 1.9 Bln words
• 1.8 Mln web pages
• Lemmatized, POS-tagged
• See special issue *English World-Wide* 36(1), 2015
Varieties for this case study

- Australia
- Ghana
- Great Britain
- Hong Kong
- India
- Jamaica
- USA
Data set

• This study: 18 Mln. words for each country
• Any form of help followed by a Vinf somewhere in the same sentence, a stop list (he, is, where, etc.)
• Decent recall and precision + manual cleaning
Frequencies
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
   • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
“Traditional” variables

- Valency of $\text{Vinf (Trans + Pass, Intrans, Clause)}$
- Helpee: Animate, Inanimate, Implicit
- Formal distance (in words) between $\text{help}$ and $\text{Vinf}$ (excluding to) (principle of complexity)
- Is there $\text{to}$ before $\text{help}$? (Horror aequi)
- Morphological form of $\text{help}$ (help, helps, helping or helped)
- Mean word length in the text: a proxy for formality
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
     • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
Ngram-based measures

• Informativity of $V_{inf}$ given the word on the left/right:
  • $-\log P(V_{inf}|\text{word left})$
  • $-\log P(V_{inf}|\text{word right})$

• Informativity of the word on the left/right given $V_{inf}$:
  • $-\log P(\text{word left}|V_{inf})$
  • $-\log P(\text{word right}|V_{inf})$

• Expectation: the greater the scores, the greater the chances of the to-$V_{inf}$

NB: the frequencies of word X with the bare and marked infinitive are added up! I.e. the bare and marked infinitives are treated as 1-gram.
Example: AU

• She also took the lead in solving the problems caused by a deconstructionist artist who got hold of an alien power wand and turned the whole city into abstract art, helped an all-female group of superheroes to battle the deadly shapeshifting Chimaera, and generally pulled her weight in the team.
  • Frequency superheroes + (to) battle = 1
  • Frequency superheroes = 22
  • Frequency (to) battle = 105
  • P(battle|superheroes) = 1/22 = 0.045
  • Informativity Vinf given word left = -log(0.045) ≈ 3.1
  • P(superheroes|battle) = 1/105 = 0.009
  • Informativity word left given Vinf = -log(0.009) ≈ 4.6
Syntactic directional measures

a) Informativity of $V_{inf}$ given help as governing verb:
   \[-\log P(V_{inf}|\text{help})\]

b) Informativity of help as governing verb given $V_{inf}$
   \[-\log P(\text{help}|V_{inf})\]

Expectation: the greater the scores, the greater the chances of the to-$V_{inf}$

NB: the frequencies of constructions with the bare and marked infinitive are added up!
Example: *get* in GB

- Frequency of *get* with *help* (as bare and to-Inf): 303
- Total frequency of *help* + *Vinf* in GB: 5950
- Total frequency of *get* in GB: 49738
- Informativity of *get* given *help* = \(-\log(303/5950) \approx 2.98\)
- Informativity of *help* given *get* = \(-\log(303/49738) \approx 5.1\)
Syntactic bidirectional measures

• Represent the mutual attraction between $V_{inf}$ and help as governing verb (cf. Levshina 2015: Ch. 10)
  • log odds ratio
  • Collostructional Strength (Gries & Stefanowitsch 2005)
  • Minimum Sensitivity (Pedersen & Bruce 2006)

• Expectation: the greater these measures, the higher the chances of the bare $V_{inf}$.
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
   • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
Technical details

• Mixed logistic models: to/bare Vinf as the response variable, random intercepts for specific Vinf and textID
• All variables centred (sum contrasts for categorical)
• 4 chains, 2000 iterations in each (50% initial discarded)
• Flat priors (as in frequentist)
• Good mixing (diagnostic plots)
• R package brms (a wrapper for Stan in C++)
Posterior probabilities of effect pro *to*-Vinf: only ‘traditional’ var.

<table>
<thead>
<tr>
<th>Variables</th>
<th>AU</th>
<th>GB</th>
<th>GH</th>
<th>HK</th>
<th>IN</th>
<th>JM</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helped vs. help</td>
<td>0.03%</td>
<td>0%</td>
<td>84.9%</td>
<td>1.4%</td>
<td>0.03%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Helping vs. help</strong></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>99.7%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Helps vs. help</td>
<td>87.9%</td>
<td>91.4%</td>
<td>84.7%</td>
<td>98.4%</td>
<td>100%</td>
<td>62.5%</td>
<td>78.1%</td>
</tr>
<tr>
<td>Tr. Vs. Intr. Vinf</td>
<td>41.3%</td>
<td>17.2%</td>
<td>63%</td>
<td>94.1%</td>
<td>84.9%</td>
<td>82.6%</td>
<td>64.1%</td>
</tr>
<tr>
<td>Clause vs. Intr. Vinf</td>
<td>54.3%</td>
<td>57.8%</td>
<td>79.6%</td>
<td>20.4%</td>
<td>13.6%</td>
<td>10.7%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Inanim. Helpee vs. anim.</td>
<td>92.4%</td>
<td>7.9%</td>
<td>15.5%</td>
<td>14.7%</td>
<td>1.1%</td>
<td>13.7%</td>
<td>86.3%</td>
</tr>
<tr>
<td><strong>Implicit Helpee vs. explicit</strong></td>
<td>96.2%</td>
<td>100%</td>
<td>99.3%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>96.2%</td>
</tr>
<tr>
<td>to help</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Ling. distance</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>99.8%</td>
</tr>
<tr>
<td>To help x ling. distance</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Mean word length</td>
<td>97.6%</td>
<td>100%</td>
<td>41.2%</td>
<td>93.4%</td>
<td><strong>0.1%</strong></td>
<td>90.6%</td>
<td>70%</td>
</tr>
</tbody>
</table>

The percentages show the posterior probabilities of a parameter having a **positive** effect on the chances of *to*-Vinf.
How to compare IT measures?

• Problem: strong correlations between some IT measures.

• Fit a model with ‘traditional’ variables + one information-theoretic for each variety (9 x 7 = 63 models)

• Compute the posterior probabilities of a positive effect wrt. to-Vinf

• LOOIC as the criterion for model comparison (predictive accuracy + parsimony)
### IT variables + ‘traditional’ (not shown)

<table>
<thead>
<tr>
<th>Ngrams</th>
<th>IT measures</th>
<th>AU</th>
<th>GB</th>
<th>GH</th>
<th>HK</th>
<th>IN</th>
<th>JM</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Info Vinf</td>
<td>70.9%</td>
<td>13.5%</td>
<td>25.3%</td>
<td>24%</td>
<td>18.4%</td>
<td>36.4%</td>
<td>59.4%</td>
</tr>
<tr>
<td></td>
<td>Word left</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Info Vinf</td>
<td>5%</td>
<td>0.3%</td>
<td>17.8%</td>
<td>6.5%</td>
<td>1.6%</td>
<td>74.4%</td>
<td>46.1%</td>
</tr>
<tr>
<td></td>
<td>Word right</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Info Word</td>
<td>100%</td>
<td>99%</td>
<td>100%</td>
<td>100%</td>
<td>98.4%</td>
<td>100%</td>
<td>88.6%</td>
</tr>
<tr>
<td></td>
<td>left</td>
<td>Vinf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Info Word</td>
<td>99.5%</td>
<td>98.7%</td>
<td>70.5%</td>
<td>42.8%</td>
<td>70.7%</td>
<td>60.5%</td>
<td>22.8%</td>
</tr>
<tr>
<td></td>
<td>right</td>
<td>Vinf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synt.</td>
<td>Info help</td>
<td>100%</td>
<td>98.1%</td>
<td>99.4%</td>
<td>100%</td>
<td>100%</td>
<td>99.7%</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td>Vinf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Info Vinf</td>
<td>77.9%</td>
<td>7.6%</td>
<td>83.8%</td>
<td>84.7%</td>
<td>33.5%</td>
<td>51.2%</td>
<td>88%</td>
</tr>
<tr>
<td></td>
<td>help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synt.</td>
<td>Log Odds</td>
<td>0.3%</td>
<td>1%</td>
<td>1.2%</td>
<td>0.05%</td>
<td>0%</td>
<td>0.9%</td>
<td>1.2%</td>
</tr>
<tr>
<td></td>
<td>Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coll.</td>
<td>Strength</td>
<td>12.7%</td>
<td>64.2%</td>
<td>2.4%</td>
<td>4.6%</td>
<td>27.2%</td>
<td>6.6%</td>
</tr>
<tr>
<td></td>
<td>Min.</td>
<td>Sensitivity</td>
<td>29.2%</td>
<td>67.4%</td>
<td>14.7%</td>
<td>20%</td>
<td>30.7%</td>
<td>26.6%</td>
</tr>
</tbody>
</table>

The percentages show the posterior probabilities of a parameter having a **positive** effect on the chances of to-Vinf.
Examples from GB

High informativity of **help**
given **Vinf**
- think
- leave
- apply
- know
- come
- describe
- publish
- tell
- use
- ....

Low informativity of **help**
given **Vinf**
- defray
- detoxify
- identify
- burnish
- regrow
- rewire
- legitimate
- demystify
- personalize
- ...

An effect of hapax legomena? No! If remove, the results remain the same.
All directional IT measures together + ‘traditional’ (not shown)

<table>
<thead>
<tr>
<th>IT measures</th>
<th>AU</th>
<th>GB</th>
<th>GH</th>
<th>HK</th>
<th>IN</th>
<th>JM</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ngrams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info Vinf</td>
<td>Word left</td>
<td>54.5%</td>
<td>30%</td>
<td>0.7%</td>
<td>10.1%</td>
<td>52.8%</td>
<td>22%</td>
</tr>
<tr>
<td>Info Vinf</td>
<td>Word right</td>
<td>13.3%</td>
<td>2.4%</td>
<td>26.7%</td>
<td>20.9%</td>
<td>15.2%</td>
<td>91.2%</td>
</tr>
<tr>
<td>Info Word left</td>
<td>Vinf</td>
<td>98.8%</td>
<td>93%</td>
<td>100%</td>
<td>99.7%</td>
<td>68.9%</td>
<td>99.6%</td>
</tr>
<tr>
<td>Info Word right</td>
<td>Vinf</td>
<td>98%</td>
<td>94.1%</td>
<td>44.9%</td>
<td>14.8%</td>
<td>30.5%</td>
<td>36.1%</td>
</tr>
<tr>
<td>Synt. direction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info help</td>
<td>Vinf</td>
<td>80.2%</td>
<td>50.9%</td>
<td>4%</td>
<td>92.6%</td>
<td>99.6%</td>
<td>82.6%</td>
</tr>
<tr>
<td>Info Vinf</td>
<td>help</td>
<td>90.8%</td>
<td>44%</td>
<td>99.7%</td>
<td>94.1%</td>
<td>44.3%</td>
<td>58.2%</td>
</tr>
</tbody>
</table>

The percentages show the posterior probabilities of a parameter having a positive effect on the chances of to-Vinf.
Outline

1. help (to) Vinf:
   • Previous research
   • Informativity hypothesis

2. GloWbE corpus

3. Variables:
   • Traditional
   • Information-theoretic

4. Bayesian mixed-effect logistic regression models

5. Discussion
Summary of the results

• There are some information-theoretic measures that play an important role in each variety.

• The estimates of other semantic and syntactic variables change little: the effects are independent.

• Overall, the higher information content, the more frequent the *to-Vinf*.

• Strikingly, it is informativity of a context given *Vinf* that matters in all varieties, rather than informativity of *Vinf* itself given the context.

• In particular, predictability of *help* as the governing verb given *Vinf* and/or that of the word on the left are important in each variety.

• Bidirectional measures do not add much in terms of predictive power to the unidirectional measures.
Discussion


• Here: longer coding after more informative contexts given the infinitive!

• Not so strange: *to* signals the reader that *Vinf* belongs together with *help* and Helpee, not with some other construction. If *Vinf* is a ‘promiscuous’ verb, it tends to be marked with *to*. If it is ‘faithful’ to *help* and Helpee, the particle will be omitted.
Thanks for your attention!

natalia.levshina@uni-leipzig.de

The slides are available at

www.natalialevshina.com/presentations.html