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

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

or

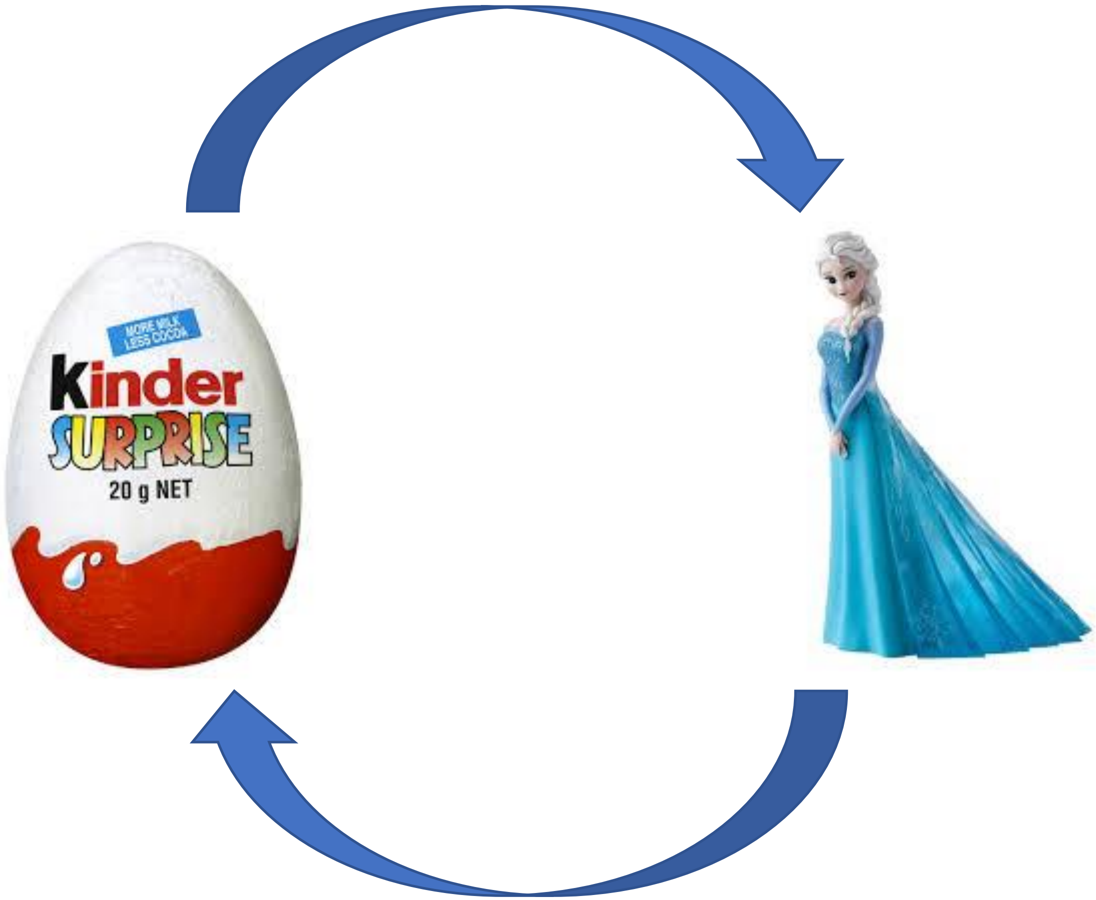
Predictability given syntactic information, e.g. Jaeger (2010): predictability of complement clauses given matrix verbs helps predict the use or omission of *that*, e.g. *I know he did it* vs. *I read that he did it.*

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 Vinf given the context? Or predictability of the context given Vinf?

Recall the intro talk…



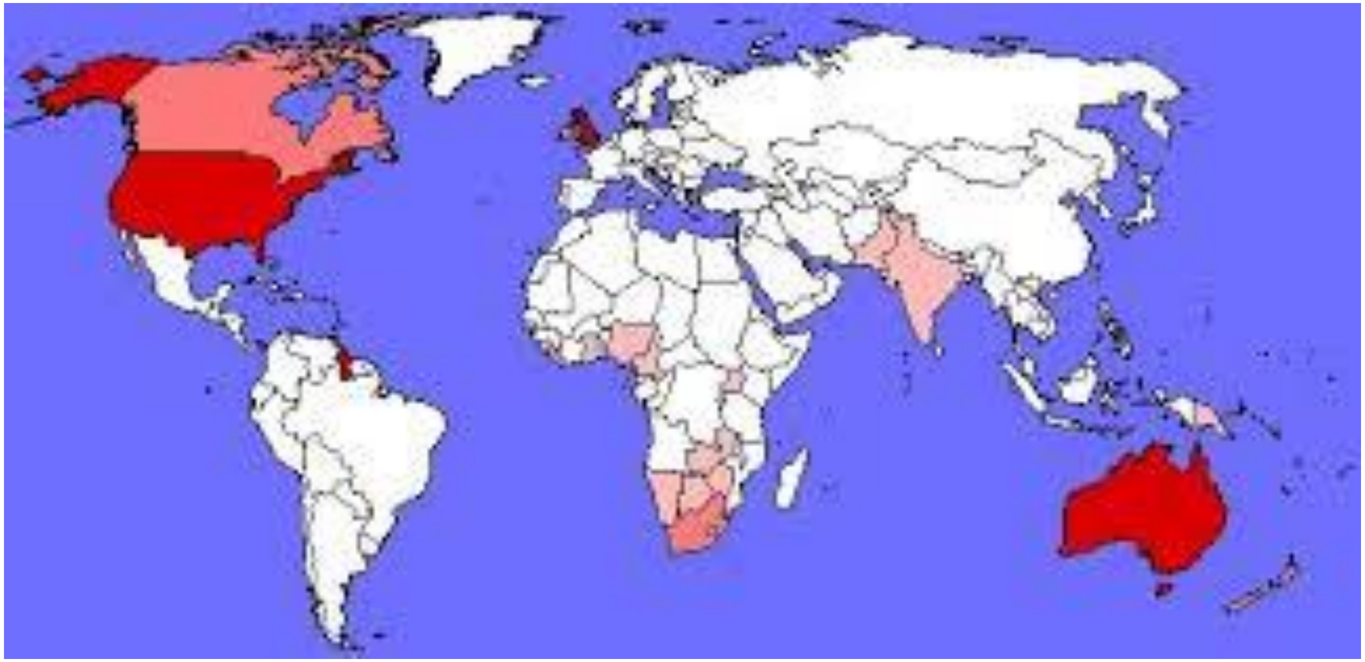
Research questions

- 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

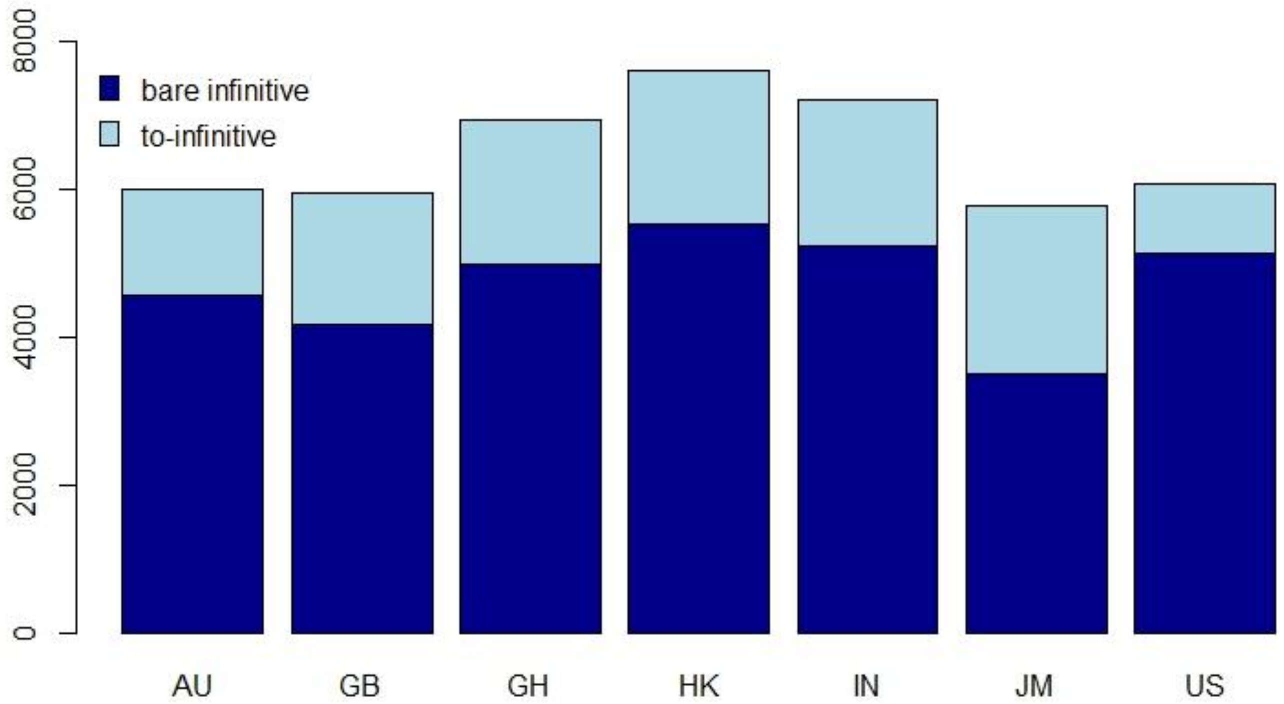
Varieties for this case study

- 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 Vinf (Trans + Pass, Intrans, Clause)
- Helpee: Animate, Inanimate, Implicit
- Formal distance (in words) between help and Vinf (excluding to) (principle of complexity)
- Is there to before help? (Horror aequi)
- Morphological form of 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 Vinf given the word on the left/right:
  - -log P(Vinf|word left)
  - -log P(Vinf|word right)
- Informativity of the word on the left/right given Vinf:
  - -log P(word left|Vinf)
  - -log P(word right|Vinf)
- Expectation: the greater the scores, the greater the chances of the to-Vinf

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 Vinf given help as governing verb:

  -log P(Vinf|help)

b) Informativity of help as governing verb given Vinf

  -log P(help|Vinf)

Expectation: the greater the scores, the greater the chances of the to-Vinf

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) ≈ 2.98
- Informativity of help given get = -log(303/49738) ≈ 5.1

Syntactic bidirectional measures

- Represent the mutual attraction between Vinf 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 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

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.

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

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)

	IT measures	AU	GB	GH	HK	IN	JM	US
Ngrams	Info Vinf | Word left	70.9%	13.5%	25.3%	24%	18.4%	36.4%	59.4%
	Info Vinf | Word right	5%	0.3%	17.8%	6.5%	1.6%	74.4%	46.1%
	Info Word left| Vinf	100%	99%	100%	100%	98.4%	100%	88.6%
	Info Word right| Vinf	99.5%	98.7%	70.5%	42.8%	70.7%	60.5%	22.8%
Synt. direction al	Info help | Vinf	100%	98.1%	99.4%	100%	100%	99.7%	99%
	Info Vinf | help	77.9%	7.6%	83.8%	84.7%	33.5%	51.2%	88%
Synt. bidirectional	Log Odds Ratio	0.3%	1%	1.2%	0.05%	0%	0.9%	1.2%
	Coll. Strength	12.7%	64.2%	2.4%	4.6%	27.2%	6.6%	6.4%
	Min. Sensitivity	29.2%	67.4%	14.7%	20%	30.7%	26.6%	31.7%

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)

	IT measures	AU	GB	GH	HK	IN	JM	US
Ngrams	Info Vinf | Word left	54.5%	30%	0.7%	10.1%	52.8%	22%	58.3%
	Info Vinf | Word right	13.3%	2.4%	26.7%	20.9%	15.2%	91.2%	64.2%
	Info Word left| Vinf	98.8%	93%	100%	99.7%	68.9%	99.6%	63.4%
	Info Word right| Vinf	98%	94.1%	44.9%	14.8%	30.5%	36.1%	12.5%
Synt. direction al	Info help | Vinf	80.2%	50.9%	4%	92.6%	99.6%	82.6%	97.2%
	Info Vinf | help	90.8%	44%	99.7%	94.1%	44.3%	58.2%	78.5%

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

- Hypothesis of UID in Levy & Jaeger (2007): maximization of uniformity of upcoming-event probabilities.
- 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