

# Mapping out particle placement around the globe: A corpus study of indigenization patterns

Jason Grafmiller

KU Leuven  
Quantitative Lexicology and Variational Linguistics



*ISLE 4*  
September 19, 2016

# Current project

"Exploring probabilistic grammar(s) in varieties of English around the world"<sup>1</sup> (2013–18, PI: B. Szmrecsanyi)

**Key questions:** To what extent do language-internal constraints on grammatical choices vary across regions?

- how stable/variable are effects of individual factors across varieties?
- how similar are regional varieties' grammars overall?
- how to explain cross-varietal differences (or lack thereof)?

---

<sup>1</sup><http://wwwling.arts.kuleuven.be/qlvl/ProbGrammarEnglish.html>

## Current project

a caveat: we focus mainly on synchronic perspective

Today I'd like to explore relevant questions from a diachronic perspective:

- do synchronic regional contrasts reflect stable nativization patterns vs. ephemeral diachronic fluctuation?<sup>2</sup>
- to what extent do synchronic regional contrasts parallel intra-variety diachronic trends?<sup>3</sup>
- how can the study of New Es (diachronic and synchronic) help us understand the nature of syntactic variation?

---

<sup>2</sup>e.g. Mair (2015)

<sup>3</sup>e.g. Wolk et al. (2013)

# Outline

1. Introduction
2. Probabilistic indigenization
3. Case study: Particle placement in 9 ICE corpora
4. Discussion & directions for future research

## Probabilistic indigenization

Process by which statistical associations among competing forms and their respective cues are reshaped by shifting usage frequencies among speakers of post-colonial varieties.<sup>4</sup>

- represents the most abstract level of structural nativization
- need not be consistent or stable (esp. in early stages)
- contrasting cues reflect the emergence of unique, region-specific grammar(s)

---

<sup>4</sup>Szmrecsanyi et al. (2016)

## Usage-based approach

- grammar is inherently variable, and variation is shaped by numerous factors
- patterns in individuals' usage lead to population-level linguistic phenomena, and speakers implicitly learn the quantitative distributions over these phenomena
- implicit statistical knowledge is reflected in users' linguistic and meta-linguistic behavior
- the nature of factors shaping variation is a current topic of debate
  - functional models assume deep-rooted (universal) cognitive biases in processing
  - purely distributional models minimize impact of such biases  
*"The grammar is the distribution"*  
(Harald Baayen, ISLE 3)

# Case study: Particle placement

- Lots of research on particle verbs<sup>5</sup>
- we focus on variable ordering of **particle** and **direct object**

## 'joined' variant (V-P-Obj)

(1) ... waiting for your lawyer to write **down the answers**. (GloWbE:NZ)

## 'split' variant (V-Obj-P)

(2) There is no need to write **the answers down** or remember them.  
(GloWbE:GB)

---

<sup>5</sup>Biberauer et al. (2014), Cappelle (2006), Diessel & Tomasello (2005), Gilquin (2014), Gries (2003, 2011), Schneider (2004), Thim (2014), Zipp & Bernaisch (2012),...

# Particle placement

- recent studies show regional contrasts in some factors, e.g. givenness and length of direct object<sup>6</sup>
  - no extant studies comparing NL varieties with ESL/EFL varieties
- stylistically sensitive
  - V-Obj-P more common in NL informal speech
  - ESL/EFL users show less register variation<sup>7</sup>
- difficult to acquire<sup>8</sup>
  - ESL/EFL learners struggle with phrasal verbs in general
  - how might this affect usage in New Englishes?

---

<sup>6</sup>Haddican & Johnson (2012)

<sup>7</sup>Gilquin (2014)

<sup>8</sup>e.g. Bronsteyn & Gustafson (2015), Diessel & Tomassello (2005), Gries (2011)

# Varieties investigated



9 varieties in ICE

- 4 NL: GB, CAN, NZ, IRE
- 5 ESL: SIN, JA, HK, PHI, IND

## Particle verb extraction

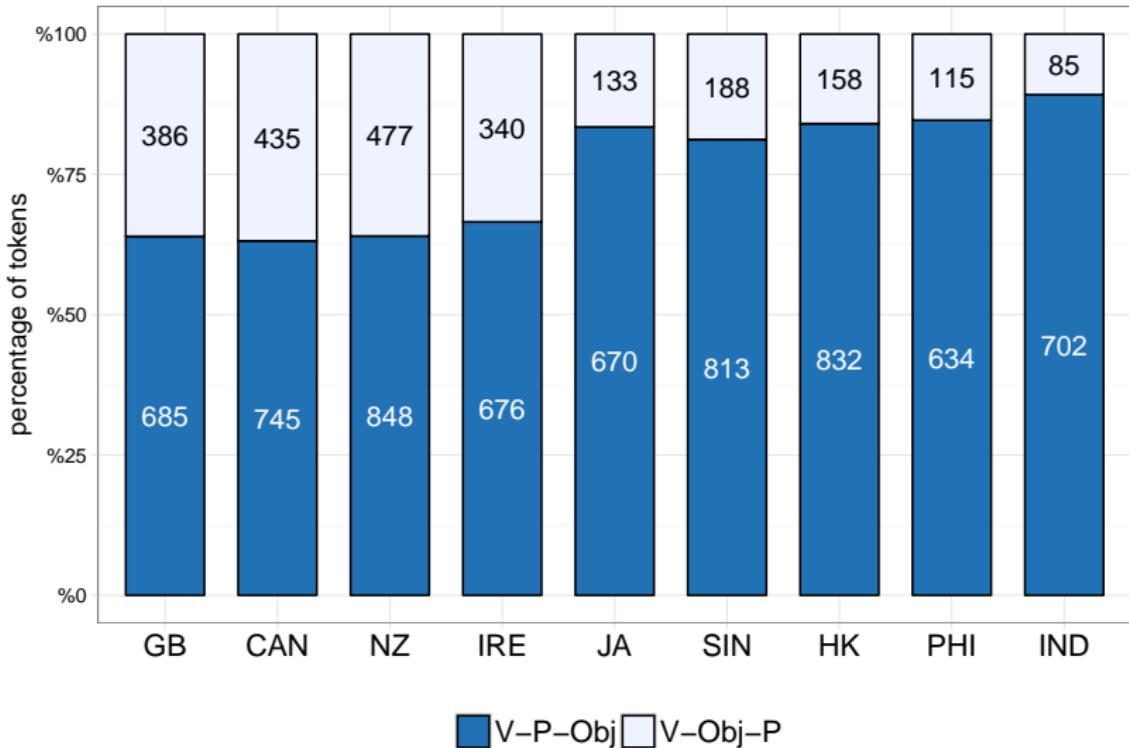
- all verbs occurring with 10 most frequent particles
  - *around, away, back, down, in, off, on, out, over, up*
- manual verification of interchangeability
  - independent check against other datasets (e.g. GloWbE, Google)
- automatic and manual annotation for numerous features
  - register
  - pronominality, length, concreteness, definiteness, givenness, text frequency of direct object
  - verb-particle association measures (e.g.  $\Delta P_{(p|v)}$ , PMI)
  - semantic idiomticity
  - presence of postmodifying PP

# Particle verb extraction

- all verbs occurring with 10 most frequent particles
  - *around, away, back, down, in, off, on, out, over, up*
- manual verification of interchangeability
  - independent check against other datasets (e.g. GloWbE)

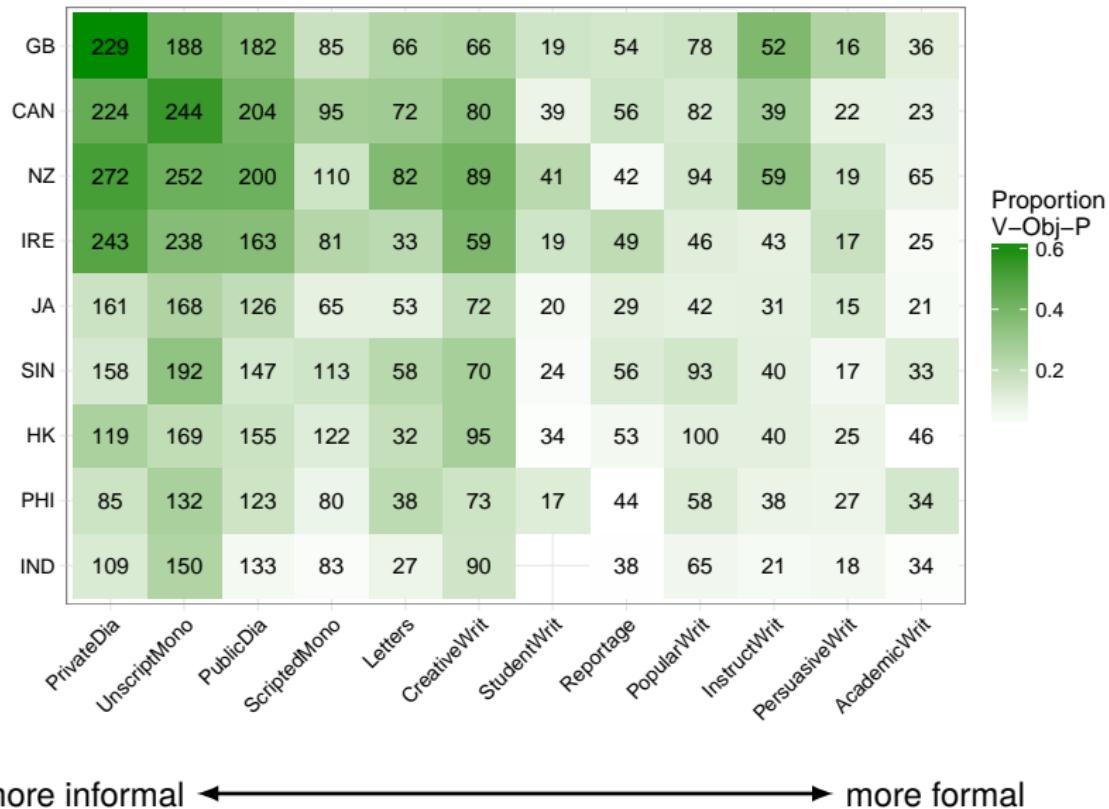
- (3) a. Resist the urge to **pull down** the shades [literal]  
b. We'll **chalk** her rude behaviour **up** to stress [idiomatic]
- (4) a. I **picked** the rock **up** off the ground  
b. I **picked up** the rock off the ground
  - presence or postmodifying PP

## Data: Lexical objects only ( $N = 8,922$ )



# Register variation

Lexical direct objects only (N = 8,922)



# Multivariate analysis

Examine the data from various angles using a number of tools

- e.g. regression, random forests, MBL, NDL<sup>9</sup>

Mainly look for...

- significant cross-varietal contrasts (interactions) in specific factors
- consistent patterns among varieties/factors
- overall cross-varietal similarities in the particle placement grammar

---

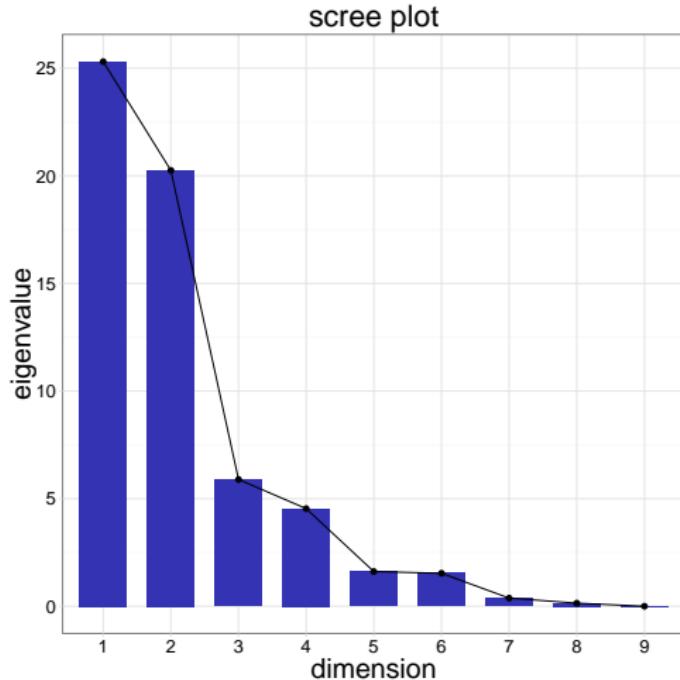
<sup>9</sup>Baayen (2011); Daelemans & van den Bosch (2009)

# Probabilistic similarity

- construct model predicting particle placement:  
variant ~ length + concreteness + ...
- fit same model to each individual variety dataset
- use model coefficients to measure distance between varieties' grammars

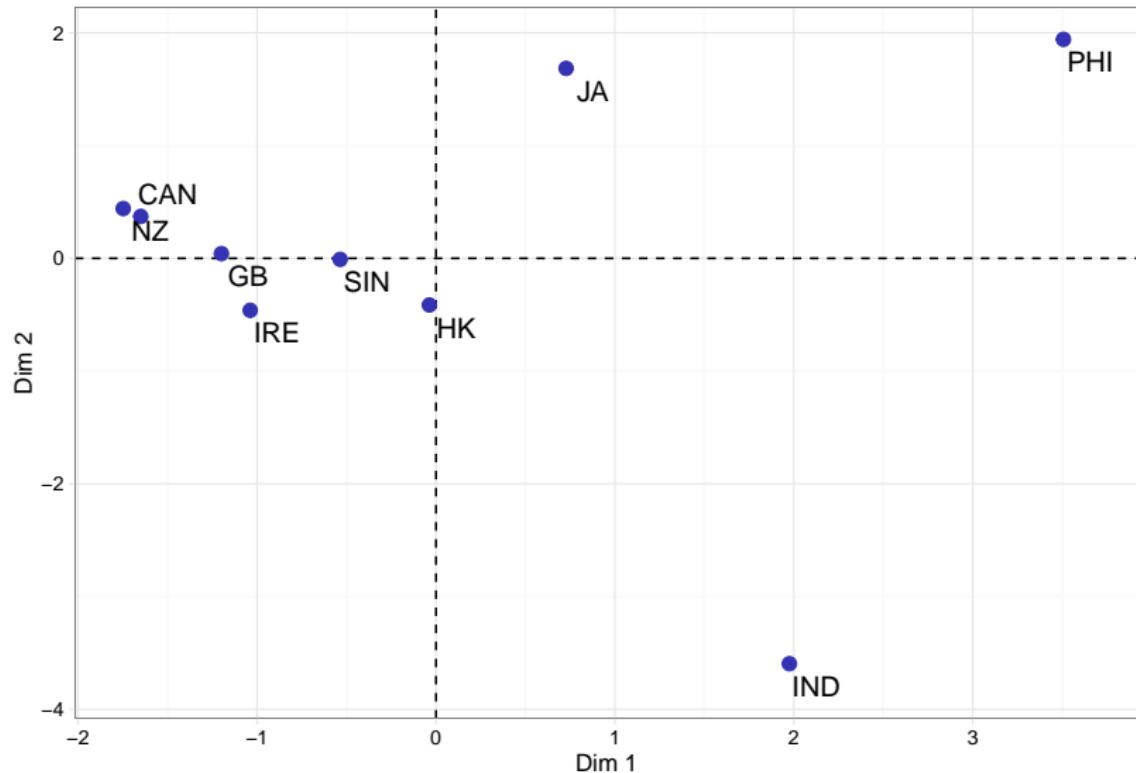
	GB	CAN	NZ	IRE	JA	SIN	HK	PHI	IND
(Intercept)	-2.34	-1.83	-1.68	-2.42	-4.55	-2.21	-2.78	-6.21	-3.51
writ.informal	-0.31	0.53	0.08	0.98	-0.03	0.33	0.41	0.88	0.11
spok.formal	0.81	1.51	1.05	1.73	0.22	-0.16	0.66	1.18	-0.35
spok.informal	1.68	1.43	1.52	1.53	0.56	0.26	0.49	1.57	1.62
DO.length	-3.29	-3.57	-3.46	-3.32	-3.63	-4.16	-4.49	-6.16	-4.76
idiomatic	-0.72	-1.11	-1.26	-0.51	0.07	-0.64	-1.39	-0.64	0.53
DO = non-conc	-0.26	-0.05	0.27	-0.4	0.49	-0.18	-0.76	0.68	-0.95
DO = given	0.99	0.8	0.89	0.89	0.77	0.07	0.83	0.26	-0.8
DO = def	1.62	1.06	1.3	0.8	1.59	1.22	1.78	2.11	1.25
DO.freq	0.37	-0.49	-0.17	0.12	-0.21	0.02	-0.56	-0.41	0.61
modifying PP	2.32	2.47	2.68	1.56	3.32	2.68	1.41	2.68	-1.25
delta P	-1.24	-1.21	-1.67	-1.52	-1.07	-2.08	-2.05	-2.27	-2.16

# Multidimensional scaling

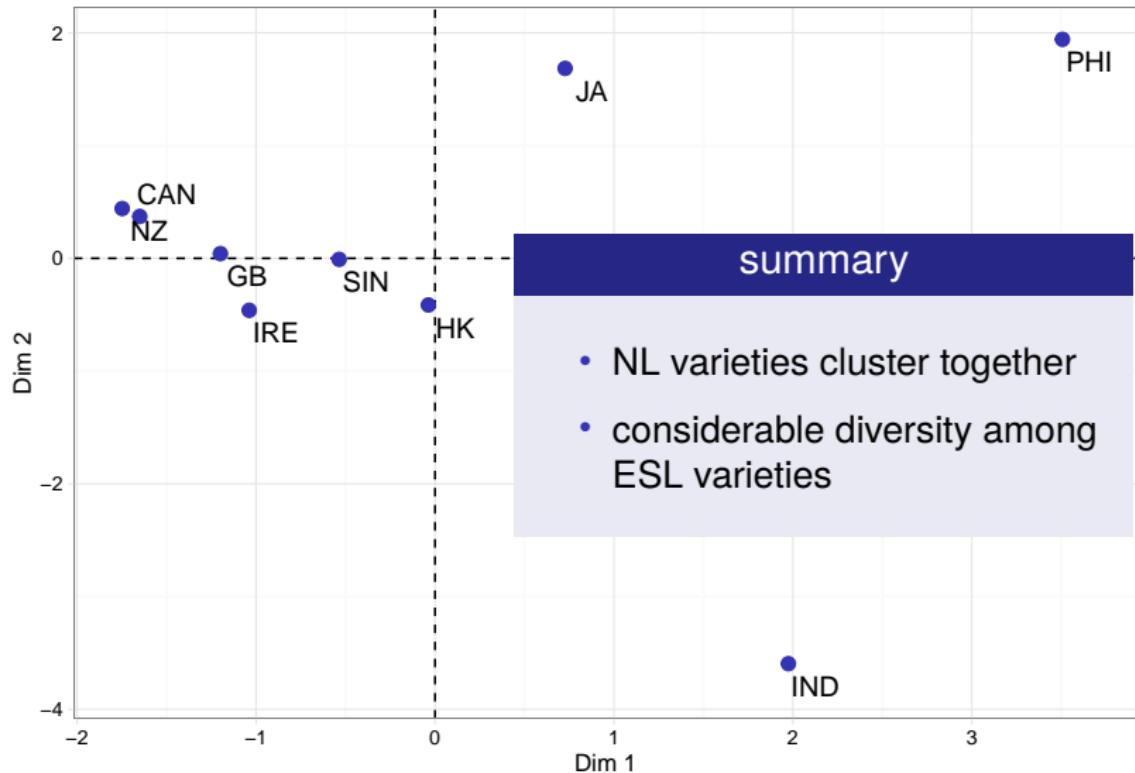


- reduce complex space to a small # of dimensions
- map top dimensions onto 2/3-dimensional surface
- the nearer the varieties on the map, the more similar their grammars

# MDS 2D similarity map



# MDS 2D similarity map



## MDS 3D similarity map (“ICE cube”)

## Interaction effects in a single model

- outlying varieties show the most/strongest contrasts (compared to GB), e.g.
  - IND ⇒ effects of object length, givenness, verb semantics, and postmodifying PP
  - PHI ⇒ effects of object length, concreteness, givenness
- more variability across registers than internal factors
  - effect of (in)formality much weaker in ESL varieties
- some consistency in interaction effects across varieties
  - few contrasts among NL varieties
  - Indian E and Philippines E most different from GB

## Interim summary

- compared to NL varieties, ESL varieties:
  - use fewer particle verbs overall
  - are less likely to use V-Obj-P variant overall
  - are less sensitive to register influences on particle placement
  - exhibit numerous differences in the strength of internal factors' effects
- NL varieties, in aggregate, exhibit markedly similar grammars of particle placement

## Discussion: Conflicting trends?

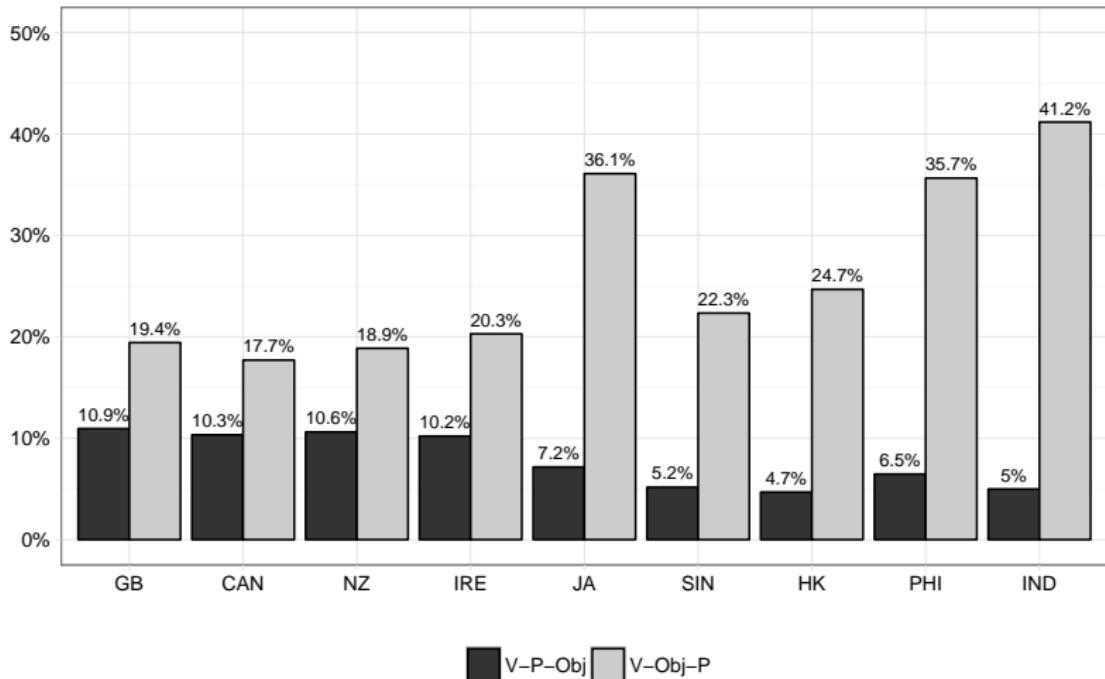
- consistent tendency for ESL varieties to use fewer V-Obj-P variants than NL varieties overall
  - just drifting in the same direction?
  - a natural part of the nativization process? → L2 acquisition?
- the factors influencing variant choice in ESL varieties seem to vary a lot
  - region-specific lexical innovation? substrate influences?
  - why don't some New E varieties end up using *more* split particle verbs?
- competing effects of L2 acquisition and other levels of structural nativization, e.g. lexical/constructional innovation

## L2 acquisition/usage

- ESL/EFL learners have difficulty with particle verbs
  - V-P-Obj predominates in English (many non-interchangeable V-Ps)
  - learners (NL and ESL) prefer semantically transparent verbs
  - learning proceeds in item-based fashion
  - limited exposure makes generalization difficult
- “over-regularization” of V-P-Obj
  - predicting interchangeability is difficult
  - learners default to more frequent V-P-Obj where NL speakers would use V-Obj-P
  - results in statistical changes to cues between variants and features

# Assymmetry in verb interchangeability

Number of interchangeable particle verb types  
as % of all particle verb types found in each variant



# Productivity in particle verbs

- (5) It's not too late to **geek out** your infant for Halloween!
- (6) How Would You **Nerd Up** The Holidays?
- (7) Here are 11 quirky accessories to **geek out** your office
- (8) Check out some awesome and very nerdy ideas for how to **nerd out** your nursery..



# Lexical innovation, style, and language change

*geek/nerd out/up X* → ‘make X geeky/nerdy’

- X is often concrete, definite (short?)
- informal, playful, like other innovative particle verbs
- innovative uses over time gradual build up to changes in associations between variants and certain linguistic cues
  - e.g. strengthened effects of concreteness, definiteness, semantic transparency

## Summing up

The broader synchronic picture reveals trends that reflect competition among different forces shaping (probabilistic) indigenization

Remaining questions...

- to what extent is nativization shaped by L2 acquisition processes/limitations?
  - can we distinguish L2 effects at the individual vs. community level?
- do contrasts in individual factors represent “random” dialect drift patterns? substrate effects?
- how are changes in Ig exposure via new technology and social media affecting usage of stylistically sensitive phenomena?

## Predictions/Hypotheses

w.r.t frequency...

- as New Es move toward endonormatively stabilized states, use of V-Obj-P variants may increase to levels similar to NL varieties

w.r.t. the constraints on placement choice... 2 possibilities

- strongly functional theories (e.g. Gries 2003, MacDonald 2013) predict convergence with NL → effects driven by processing demands
- strongly distributional theories (e.g. Baayen & Ramscar 2015) predict stable divergence → effects driven by localized lexical/constructional innovations

## Where to go from here?

- more real-time diachronic studies of particle placement in NL and ESL varieties in later stages of development (e.g. SIN, HK)
- explore diachronic developments in New Es through lens of L2 acquisition/processing
- Comparison of patterns in contemporary EFL varieties with earlier stages of ESL varieties
- investigation of particle verb productivity and socio-stylistic variation new media (e.g. Twitter, IM)

# Thank you!

Contact:

jason.grafmiller@kuleuven.be

Special thanks to Christy Tanh Linh Ha for help with extraction and annotation, Melanie Röthlisberger for help with visualization.

# References I

- Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada* 11(2). 295–328.  
doi:10.1590/S1984-63982011000200003.
- Baayen, R. Harald & Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 100–120. Berlin: De Gruyter.
- Biberauer, Theresa, Liliane Haegeman & Ans van Kemenade. 2014. Putting our heads together: Towards a syntax of particles. *Studia Linguistica* 68(1). 1–15.  
doi:10.1111/stul.12021.
- Bronshteyn, Karen & Tom Gustafson. 2015. The Acquisition of Phrasal Verbs in L2 English: A Literature Review. *Linguistic Portfolios* 4(1).
- Cappelle, Bert. 2006. *Particle placement and the case for 'allostructions'*. Published: In *Constructions special volume 1. Constructions all over: Case studies and theoretical implications*.
- Daelemans, Walter & Antal van den Bosch. 2009. *Memory-based language processing*. Cambridge; New York: Cambridge Univ. Press.
- Diessel, Holger & Michael Tomasello. 2005. Particle placement in early child language: A multifactorial analysis. *Corpus Linguistics and Linguistic Theory* 1. 89–112.
- Gilquin, Gaëtanelle. 2014. The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory* 11(1). 51–88. doi:10.1515/cllt-2014-0005.

## References II

- Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum Press.
- Gries, Stefan Th. 2011. Acquiring particle placement in English: A corpus-based perspective. In Pilar Guerrero Medina (ed.), *Morphosyntactic alternations in English: Functional and cognitive perspectives*, 235–263. London: Equinox.
- Haddican, Bill & Daniel Ezra Johnson. 2012. Effects on the particle verb alternation across English dialects. In *University of Pennsylvania Working Papers in Linguistics* 18, 31–40. University of Pennsylvania.
- Mair, Christian. 2015. Cross-variety diachronic drifts and ephemeral regional contrasts: An analysis of modality in the extended Brown family of corpora and what it can tell us about the New Englishes. In Peter Collins (ed.), *Studies in Corpus Linguistics*, vol. 67, 119–146. Amsterdam: John Benjamins Publishing Company.
- Schneider, Edgar. 2004. How to trace structural nativization: Particle verbs in world Englishes. *World Englishes* 23(2). 227–249.
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137. doi:10.1075/eww.37.2.01szm.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsányi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30. 382–419.