Epicentral Configurations in South Asian Englishes

Tobias Bernaisch
(Justus Liebig University Giessen)

(paper co-authored with
Stefan Th. Gries
(University of California, Santa Barbara;
 Justus Liebig University Giessen))

1 Theoretical Background

1.1 The dative alternation

1.2 South Asia & South Asian Englishes

1.3 Linguistic epicentres

2 Methodology

2.1 Corpus data

2.2 Data coding

3 Case Studies

3.1 Sources of structural nativisation in South Asian Englishes

3.2 Identifying the South Asian linguistic epicentre

4 Discussion

1.1 The dative alternation

Alternation between the

double-object construction (e.g.   *John gave Mary a book.*)
and the prepositional dative (e.g.   *John gave a book to Mary.*)

Factors influencing the choice of one variant over another (cf. e.g. Gries 2003; Bresnan & Hay 2008; Schilk et al. 2013; Bernaisch et al. 2014):

• animacy of patient/recipient

• discourse accessibility of patient/recipient

• length of patient/recipient

• pronominality of patient/recipient

• semantics of patient

• variety

1.2 South Asia & South Asian Englishes
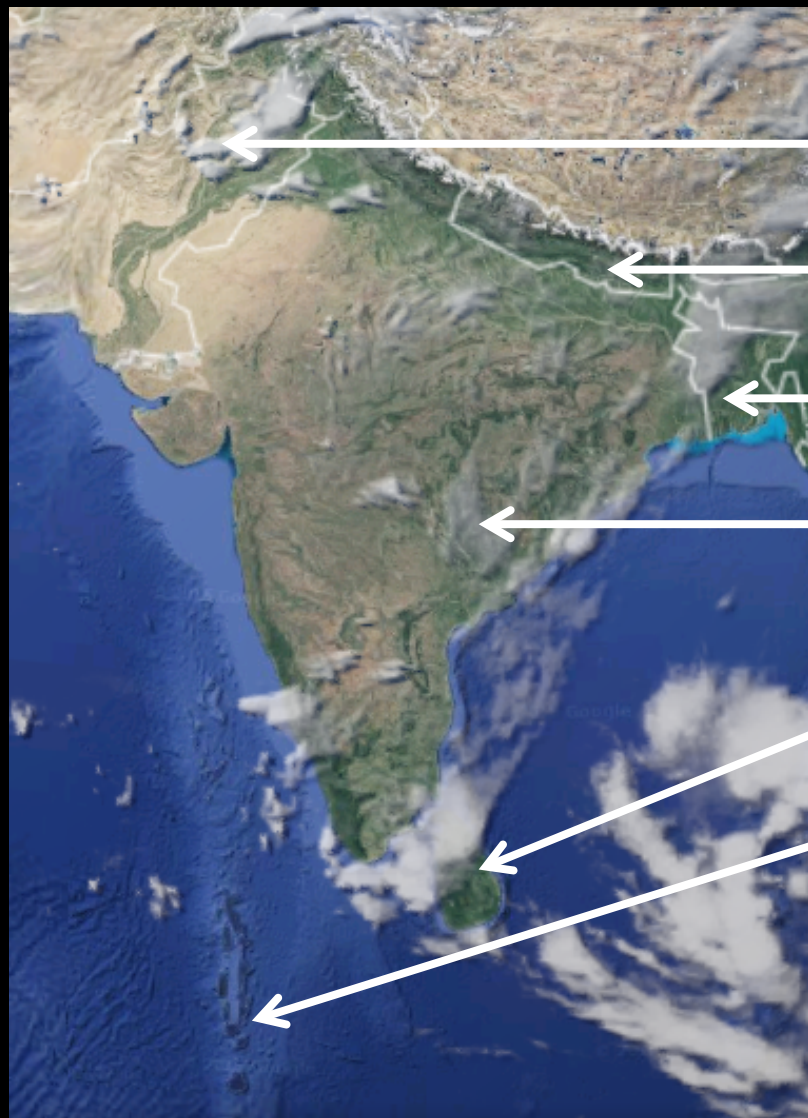
Pakistani English

Nepali English

Bangladeshi English

Indian English

Sri Lankan English

Maldivian English

(taken from Google Maps)

(taken from Google Maps)

"[A] variety can be regarded as a potential epicentre if it shows endonormative stabilization (i.e. widespread use, general acceptance and codification of the local norms of English) [...] on the one hand, and the potential to serve as a model of English for (neighbouring?) countries on the other hand." (Hundt 2013: 185)

Epicentre research has so far mainly relied on inferring interpretations from "degrees of similarity between a specific dominant variety on the one hand (i.e. British English or Indian English) and peripheral varieties on the other (e.g. Sri Lankan English and Pakistani English)" (Hoffmann et al. 2011: 261).

(taken from Google Maps)

Structural similarities across South Asian Englishes (SAEs) supporting the status of Indian English as an epicentre:

- Hoffmann et al. (2011) on light-verb constructions
- Bernaisch & Lange (2012) on presentational *itself*

Structural differences across SAEs not supporting the status of Indian English as an epicentre:

- Hundt et al. (2012) on the hypothetical subjunctive
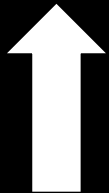- Koch & Bernaisch (2013) on new ditransitives

"[A] variety can be regarded as a potential epicentre if it shows endonormative stabilization (i.e. widespread use, general acceptance and codification of the local norms of English) [...] on the one hand, and **the potential to serve as a model of English for (neighbouring?) countries on the other hand**." (Hundt 2013: 185)

<u>Dative alternation</u>

Surface-structure choices

Double-object construction
vs. prepositional dative

⬆

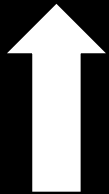⬆

Norms constituting
(variety-specific) models
and guiding surface-structure choices

Norms guiding the choice of either the double-object construction or the prepositional dative

"[A] variety can be regarded as a potential epicentre if it shows endonormative stabilization (i.e. widespread use, general acceptance and codification of the local norms of English) [...] on the one hand, and **the potential to serve as a model of English for (neighbouring?) countries on the other hand**." (Hundt 2013: 185)

Surface-structure choices

↑

Norms constituting
(variety-specific) models
and guiding surface-structure choices

<u>Dative alternation</u>

Double-object construction

He gave her a book.

He gave his daughter the freedom to come home late.

↑

Factors:
recipient = pronominal

↑

Factors:
recipient   ≤ 5 words
recipient   = non-pronominal
patient   > 3 words
patient   = abstract
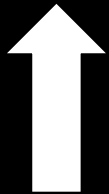
"[A] variety can be regarded as a potential epicentre if it shows endonormative stabilization (i.e. widespread use, general acceptance and codification of the local norms of English) [...] on the one hand, and **the potential to serve as a model of English for (neighbouring?) countries on the other hand.**" (Hundt 2013: 185)

Surface-structure choices   Epicentre identification

Norms constituting
(variety-specific) models
and guiding surface-structure choices

2.1 Corpus data

South Asian Varieties of English (SAVE) Corpus featuring six national components with 3m words of newspaper language per variety (cf. Bernaisch et al. 2011)

News section from the British National Corpus for British English (BrE) reference data

Variety	Sources	URLs
Bangladeshi English	Daily Star New Age	http://www.thedailystar.net http://www.newagebd.com
Indian English	The Statesman The Times of India	http://www.thestatesman.net http://timesofindia.indiatimes.com
Maldivian English	Dhivehi Observer Minivan News	http://www.dhivehiobserver.com http://www.minivannews.com
Nepali English	Nepali Times The Himalayan Times	http://www.nepalitimes.com http://www.thehimalayantimes.com
Pakistani English	Daily Times Dawn	http://www.dailytimes.com.pk http://www.dawn.com
Sri Lankan English	Daily Mirror Daily News	http://www.dailymirror.lk http://www.dailynews.lk
British English	News section of the British National Corpus (BNC)	

2.2 Data coding

1381 examples with GIVE were annotated according to the following variables

Variable	Description	Variants
VARIETY	the variety from which the example was taken	
PAPER	the newspaper from which the example was taken	
LOGLENDIFF	logged difference between length of recipient and length of patient	[numeric value]
RECANIMACY & PATANIMACY	animacy of recipient and patient	*animate* vs. *inanimate*
RECACCESSIBILITY & PATACCESSIBILITY	discourse accessibility of recipient and patient	*given* vs. *new*
RECPRONOMINALITY & PATPRONOMINALITY	realisation of recipient and patient in pronominal or non-pronominal form	*pronoun* vs. *np*
PATSEMANTICS	semantic class of patient	*abstract* vs. *concrete* vs. *informational*
TRANSITIVITY	verb-complementational pattern of GIVE	*ditransitive* vs. *prepositional dative*

- new regression-based approach from the domain of learner corpus research called MuPDAR (<u>M</u>ultifactorial <u>P</u>rediction and <u>D</u>eviation <u>A</u>nalysis using <u>R</u>egression, see Gries & Deshors 2014):

1. generate a concordance of the phenomenon of interest from NS and NNS data and annotate for predictors P_{1-n} known/hypothesized to affect the phenomenon;
2. fit a regression R_1 to model the phenomenon as a function of P_{1-n} in the NS data only and check R_1's classification accuracy;
3. if R_1's classification accuracy is good, apply R_1 to the NNS data to, for each case, get a prediction of 'what a NS would have done here';
4. compare whether the NNS made the predicted NS choices and fit a regression R_2 to model where and how much the NNS made non-nativelike/non-idiomatic choices.

- extension of MuPDAR to the comparison of BrE (as the 'target' variety, NS in the above) to indigenised varieties (as the 'learner' varieties, NNS above): what factors are responsible for indigenised-variety speakers making non-BrE choices?

- with the annotated predictors on the BrE data, R_1 is created and its classification accuracy is evaluated; crucially, R_1 is a mixed-effects model taking the relatedness of data points from the two BNC parts into consideration;

- R_1 is applied to the indigenised-variety speakers to predict BrE speakers' choices and a variable called VARIETYSPECIFICITY stating whether non-BrE speakers made BrE choices or not is created;

- with the annotated predictors on the indigenised-variety data, R_2 is created for VARIETYSPECIFICITY; crucially, R_2 is a mixed-effects model taking the hierarchical structure of the corpus data into consideration (VARIETY/ NEWSPAPER, see Gries 2015);

- the degree of how much the non-BrE speakers' choices differed from those the BrE speakers made was also created.

Results

- R_1 resulted in a very good classification accuracy of 93.7% on the BrE data (*** better than chance)
- R_1 resulted in a good prediction accuracy of 77% on the non-BrE data (*** better than chance)
- VARIETYSPECIFICITY values were computed:
  - if the SAE speaker made the BrE choice, VARIETYSPECIFICITY = 0
  - if the SAE speaker did not make the BrE choice, VARIETYSPECIFICITY = 0.5 − predicted probability of prepositional dative
    - thus, if VARSPEC > 0, SAE user used a prep. dative, but a BrE speaker would have chosen a ditransitive
    - thus, if VARSPEC < 0, SAE user used a ditransitive, but a BrE speaker would have chosen a prep. dative
- R_2 resulted in a good classification accuracy of 77.2% (*** better than chance)

3.1 Case study 1: Identifying factors of structural nativisation in SAEs

- SAE speakers typically make BrE-like choices: most points are around *y*=0;
- the more the length difference gives a clue (i.e, the further *x* is from 0), the more BrE-like their choices are;
- when the patient is given, SAE speakers make non-BrE choices equally much (see ×);
- when the patient is new, they are much more likely to choose non-BrE-like preposi-tional datives (see ×);
- thus, compared to BrE, the strength of the cue 'new patient' is stronger for prep. datives in SAEs.

ditr.: She gave $him_{REC/given}$ a $book_{PAT/new}$.
prep. dat.: She gave the $ticket_{PAT/given}$ to a $man_{REC/new}$.
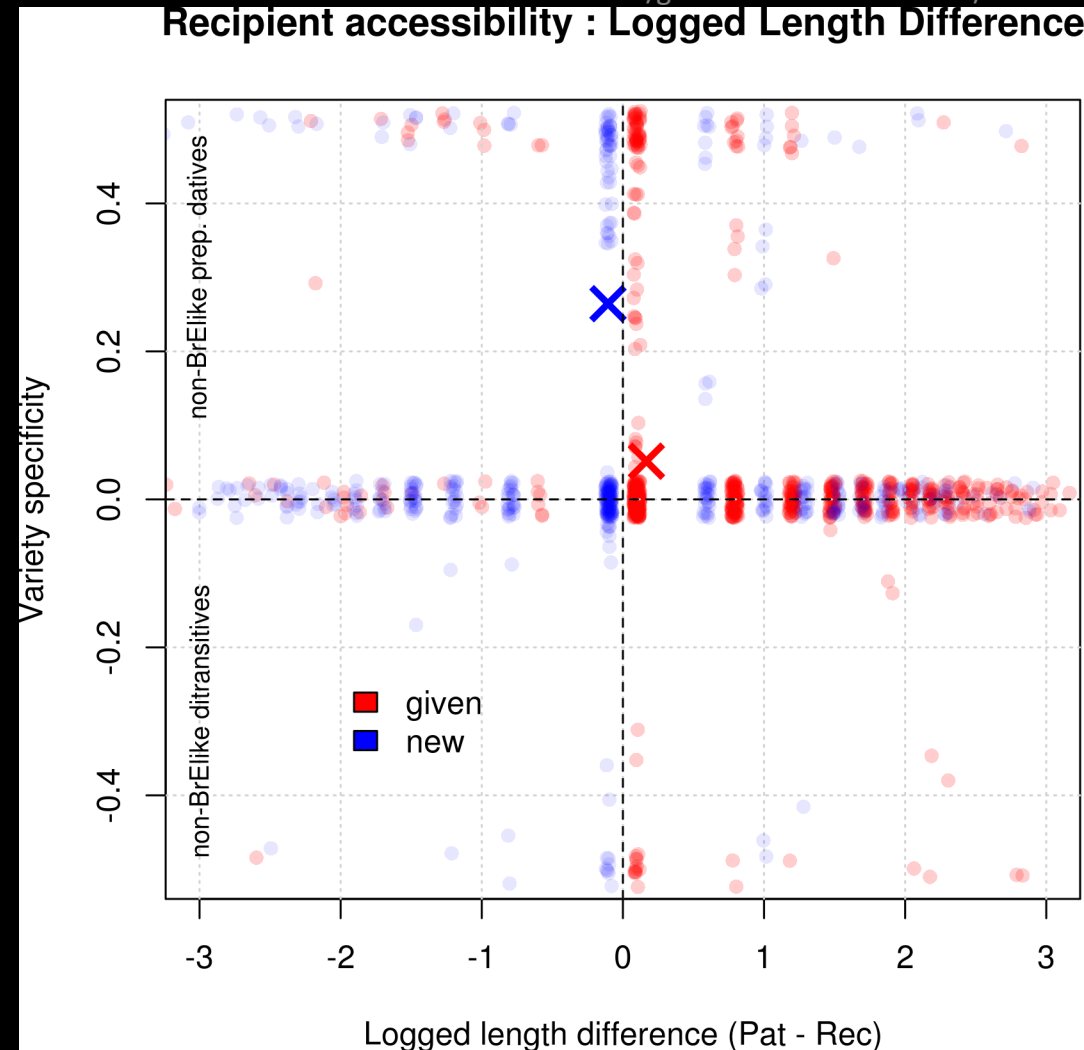
Patient accessibility : Logged Length Difference
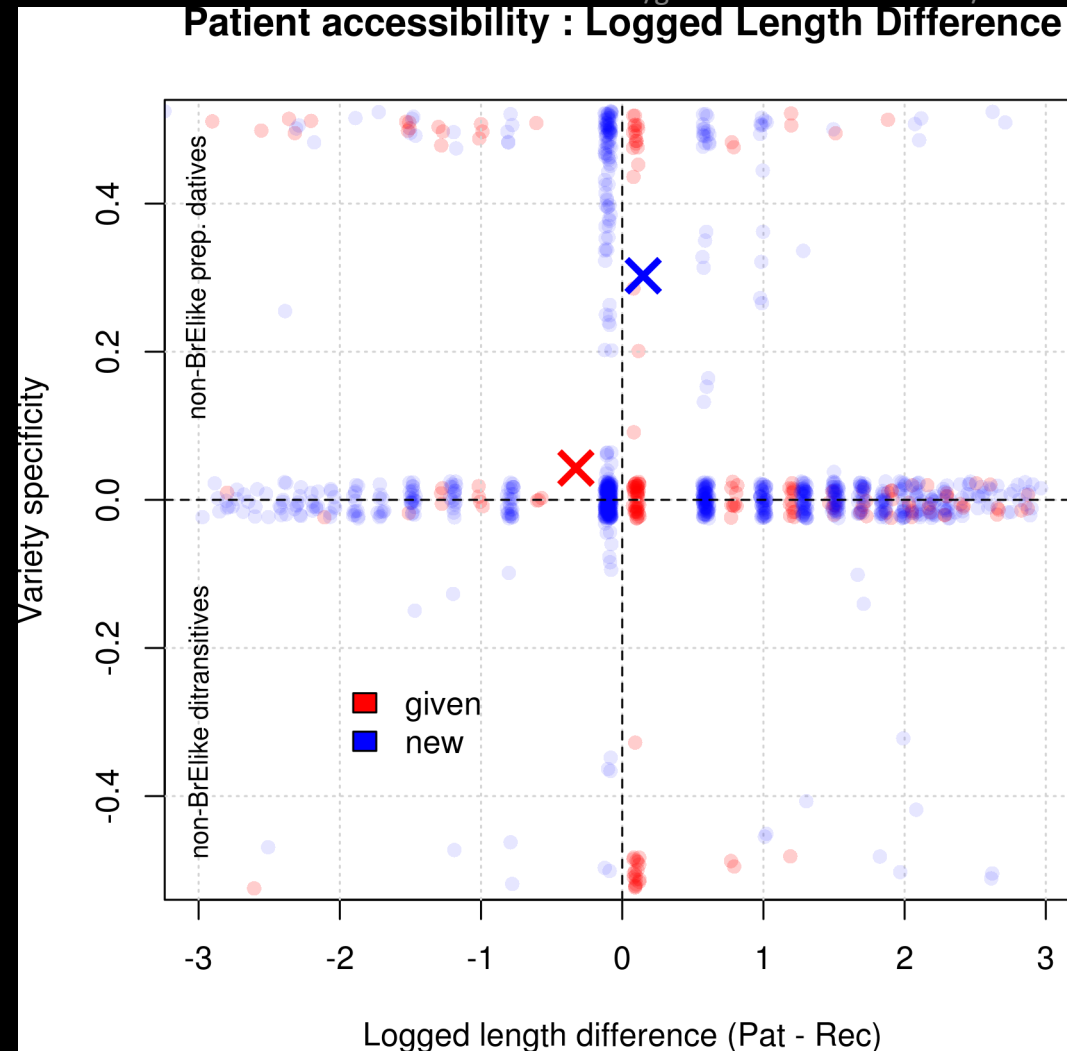
3.1 Case study 1: Identifying factors of structural nativisation in SAEs

- SAE speakers typically make BrE-like choices: most points are around *y*=0;
- the more the length difference gives a clue (i.e, the further *x* is from 0), the more BrE-like their choices are;
- when the recipient is given, SAE speakers make non-BrE choices equally much (see ×);
- when the recipient is new, they are much more likely to choose non-BrE-like preposi-tional datives (see ×) – it seems in fact as if the cue 'new recipient' for prep.dat. is stronger for SAEs than for BrE.

ditr.: She gave him$_{REC/given}$ a book$_{PAT/new}$.
prep. dat.: She gave the ticket$_{PAT/given}$ to a man$_{REC/new}$.

Recipient accessibility : Logged Length Difference

non-BrElike prep. datives

non-BrElike ditransitives

Variety specificity

■ given
■ new

Logged length difference (Pat - Rec)

3.1 Case study 1: Identifying factors of structural nativisation in SAEs

ditr.: She gave $him_{REC/given}$ a $book_{PAT/new}$.
prep. dat.: She gave the $ticket_{PAT/given}$ to a $man_{REC/new}$.

- SAE speakers typically make BrE-like choices: most points are around $y=0$;
- when the recipient is a pronoun, then the SAE choices are BrE-like (see ●), esp. when the patient is new;
- when the recipient is lexical, SAE speakers are more likely than BrE speakers to use prepositional datives (see ●).

Patient accessibility : Recipient pronominality

3.1 Case study 1: Identifying factors of structural nativisation in SAEs
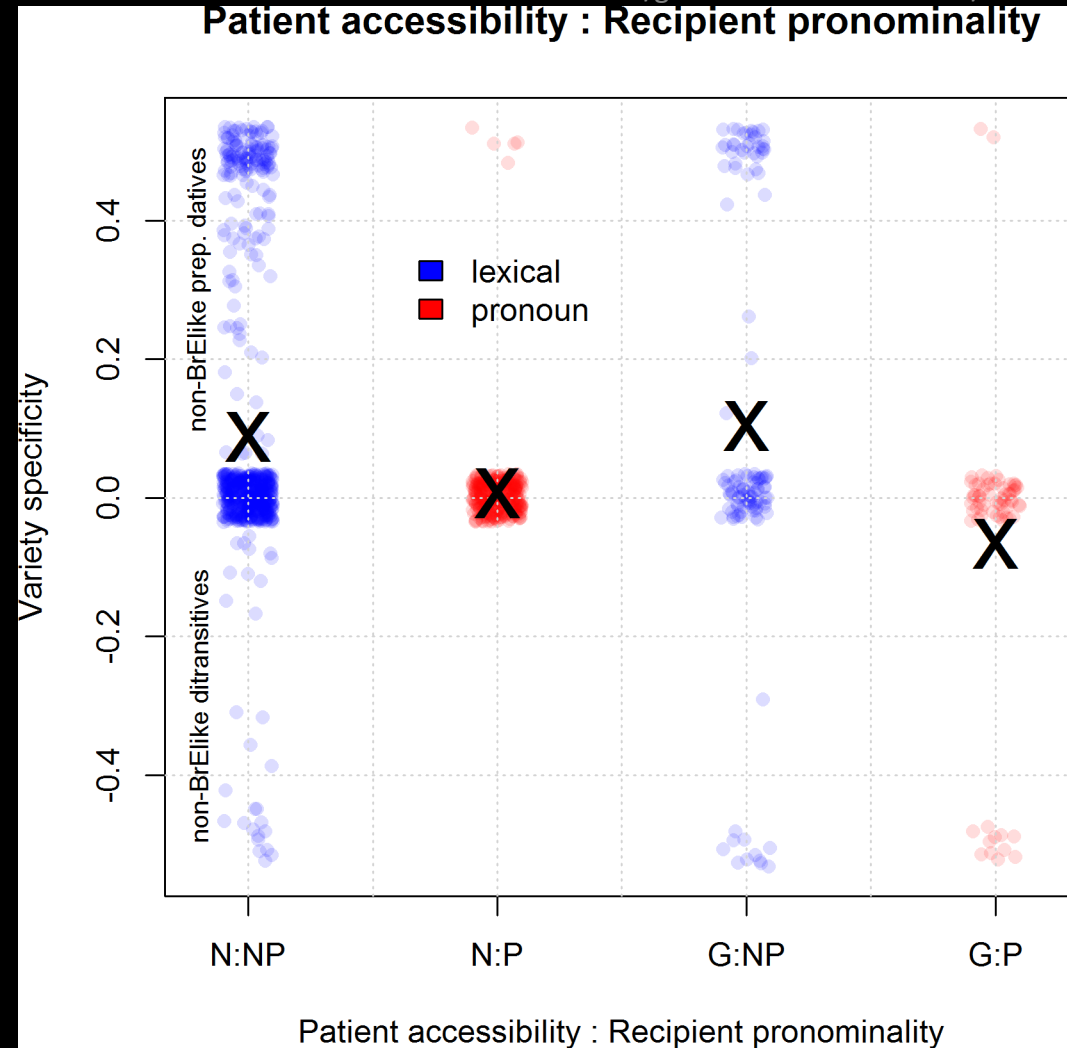
- Note: some SAVE components (Ind, Pak, SL) are much more homogeneous than others (Ban, Mal, Nep)
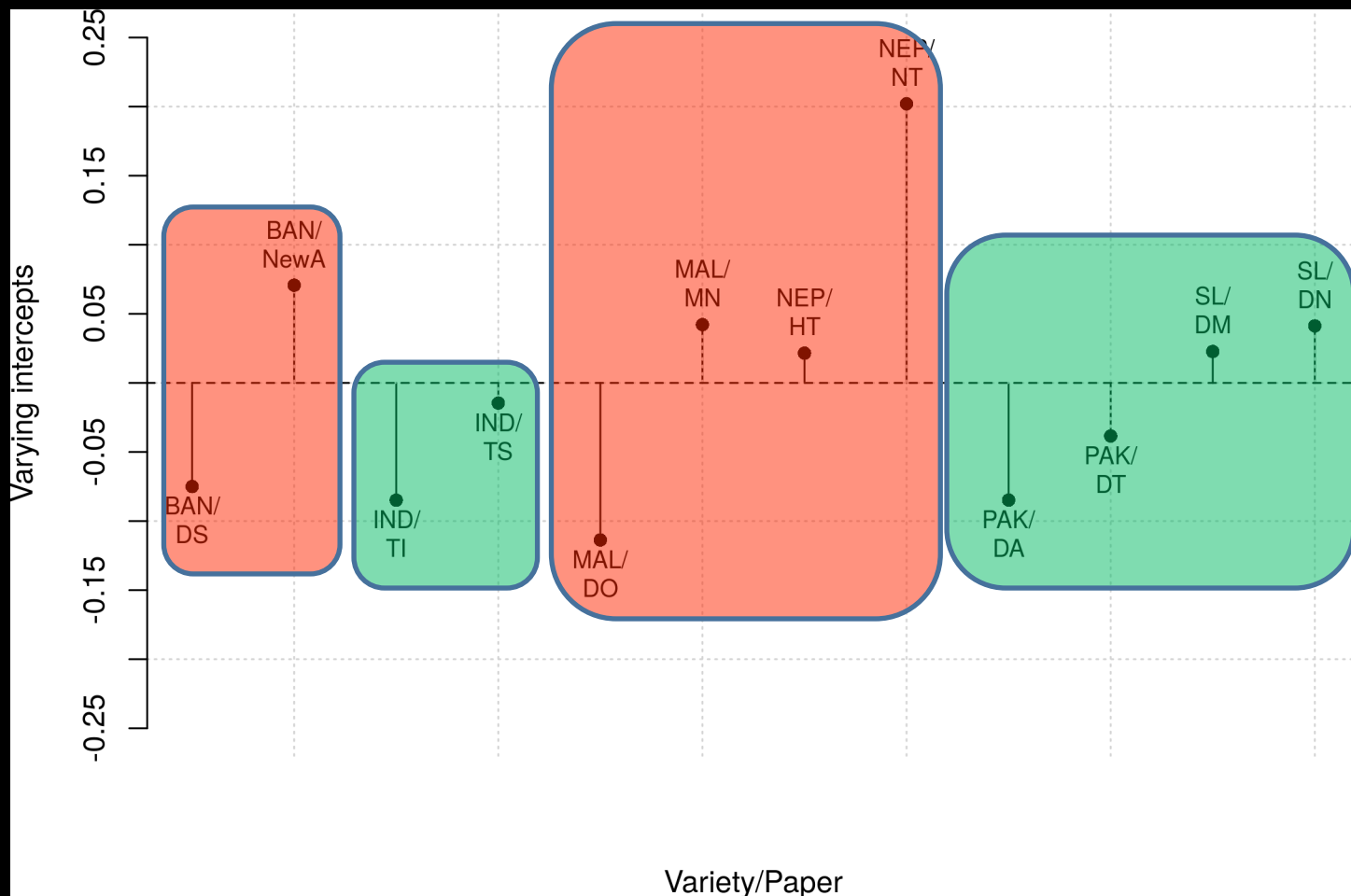
Following Hundt (2013: 185), epicentres have "the potential to serve as a model of English for (neighbouring?) countries".

Proposed operationalisation: model = set of norms governing structural choices; thus, the similarity of the norms of varieties to those of an assumed epicentre will reflect how likely the assumed epicentre *is* an epicentre.
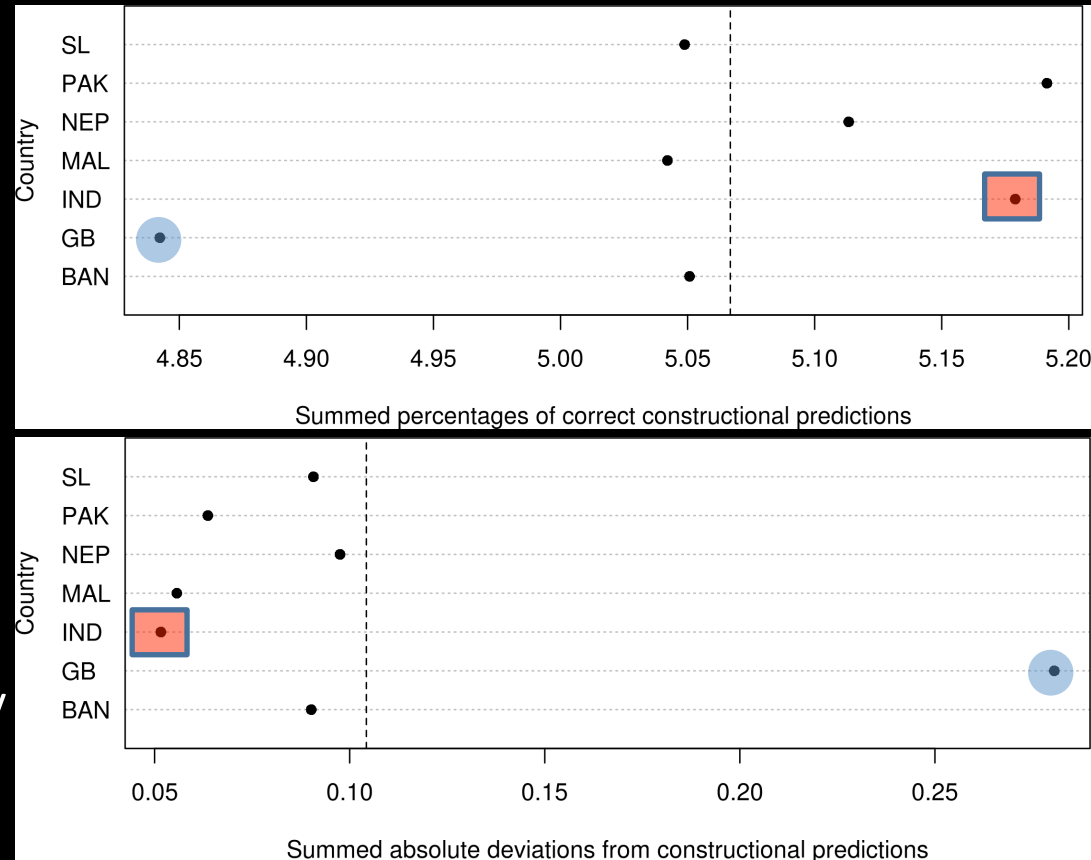
The proposed method is a bottom-up extension of MuPDAR:
- each of the varieties studied here will be assumed to be the epicentre, i.e. yield R_1-based predictions for all other varieties:
  - coarse-grained approach: %s of structural choices as predicted by the assumed epicentre;
  - fine-grained approach: absolute deviations from all varieties' users' choices from the assumed epicentre;
- then, each of these statistics was summed up for each assumed epicentre and the sums were plotted.

- **BrE** behaves quite differently from the (cluster of) SAE varieties
  - in the coarse-grained data
  - in the fine-grained data
- in combination, the results point to **IndE** as the epicentre:
  - in the coarse-grained plot, IndE is narrowly bested by PakE, but ...
  - in the fine-grained plot, IndE is the variety from which the others are predicted best (w/ the smallest sum of devia-tions).

Summed percentages of correct constructional predictions

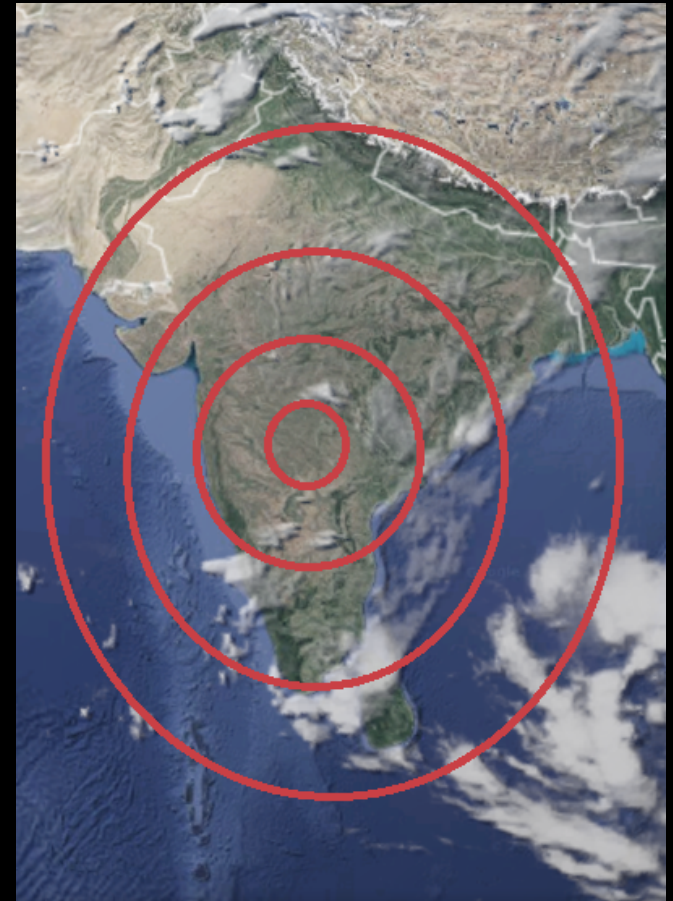Summed absolute deviations from constructional predictions

Indian English as the linguistic epicentre for South Asian Englishes?

Criteria for a linguistic epicentre:

- endonormative stabilisation?   yes
- model for other varieties in its vicinity?   yes

Desiderata:

- comparable studies of
  other phenomena/alternations
- diachronic data

- 'what would a speaker of a/the historical input variety have done' in the structural situation in which the ESL speaker finds herself or himself?

- MuPDAR can be used exploratively to examine potential epicentral configurations

- the hierarchical structure of the corpus data is taken into account by using multi-level modeling in the regression modeling

References

Bernaisch, Tobias and Claudia Lange. 2012. "The typology of focus marking in South Asian Englishes". *Indian Linguistics* 73(1–4): 1–18.

Bernaisch, Tobias, Christopher Koch, Joybrato Mukherjee and Marco Schilk. 2011. *Manual for the* South Asian Varieties of English (SAVE) Corpus*: Compilation, Cleanup Process, and Details on the Individual Components*. Giessen: Justus Liebig University.

Bernaisch, Tobias, Stefan Th. Gries and Joybrato Mukherjee. 2014. "The dative alternation in South Asian English(es): modelling predictors and predicting prototypes". *English World-Wide* 35(1): 7–31.

Bresnan, Joan and Jennifer Hay. 2008. "Gradient grammar: an effect of animacy on the syntax of *give* in New Zealand and American English". *Lingua* 118: 245–259.

Gries, Stefan Th. 2003. "Towards a corpus-based identification of prototypical instances of constructions". *Annual Review of Cognitive Linguistics* 1: 1–27.

Gries, Stefan Th. 2015. The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95-125.

Gries, Stefan Th. and Sandra C. Deshors. 2014. Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora* 9(1). 109–136.

Hoffmann, Sebastian, Marianne Hundt and Joybrato Mukherjee. 2011. "Indian English – an emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes". *Anglia* 129(3–4): 258–280.

Hundt, Marianne. 2013. "The diversification of English: old, new and emerging epicentres". In Daniel Schreier and Marianne Hundt, eds. *English as a Contact Language*. Cambridge: Cambridge University Press, 182–203.

Hundt, Marianne, Sebastian Hoffmann and Joybrato Mukherjee. 2012. "The hypothetical subjunctive in South Asian Englishes: local developments in the use of a global construction". *English World-Wide* 33(2): 147–164.

References

Koch, Christopher and Tobias Bernaisch. 2013. "Verb complementation in South Asian English(es): the range and frequency of 'new' ditransitives". In Gisle Andersen and Kristin Bech, eds. *English Corpus Linguistics: Variation in Time, Space and Genre – Selected Papers from ICAME 32*. Amsterdam: Rodopi, 69–89

Schilk, Marco, Joybrato Mukherjee, Christopher F.H. Nam and Sach Mukherjee. 2013. "Complementation of ditransitive verbs in South Asian Englishes: a multifactorial analysis". *Corpus Linguistics and Linguistic Theory* 9(2): 187–225.