University of Zurich[UZH]

English Department URPP Language and Space Universidade Vigo

Predicting Voice Alternation Across Academic Englishes

Marianne Hundt (Zürich) and Elena Seoane (Vigo)

Workshop *Probabilistic variation across dialects and varieties*

Leuven, 4-5 April 2016

Introduction

Aim: model the local use of a global linguistic variant in a specialized text type.

- use of *be*-passives in academic discourse in varieties of
  - English as a first language
  - English as an institutionalised second-language

ICE corpora, syntactically annotated at University of Zürich

- use of *be*-passives vs. transitive active constructions

Introduction

Research questions of the workshop addressed

- What is the extent to which varieties of English share a probabilistic grammar that predicts variation patterns across different varieties?
- How do we evaluate overall probabilistic similarity between varieties?

Introduction

- Voice alternation and probabilistic indigenization?

  "… the process whereby stochastic patterns of internal linguistic variation are reshaped by shifting usage frequencies in speakers of post-colonial varieties. To the extent that patterns of variation in a new variety A, e.g. the probability of item x in context y, can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A. These patterns need not be consistent or stable ..., but they nonetheless reflect the emergence of a unique, region-specific grammar." (Szmrecsanyi et al., fc. 2016)

Background

Previous research: passives across time and space

- Passives decline, especially in academic English
- Regional difference: AmE leading the change (Leech et al. 2009)
- Little research on varieties other than AmE and BrE (cf. Biewer 2007)

Hundt at al. (2016) – 15 varieties of English:

- Diachronic change: confirmed
- Regional difference: confirmed

BUT

Previous research: passives across time and space

regression analysis shows – most important factor is register variation

- Difference between subdisciplines: decrease more pronounced in natural sciences and technology than humanities and social sciences (Hundt et al. 2016)

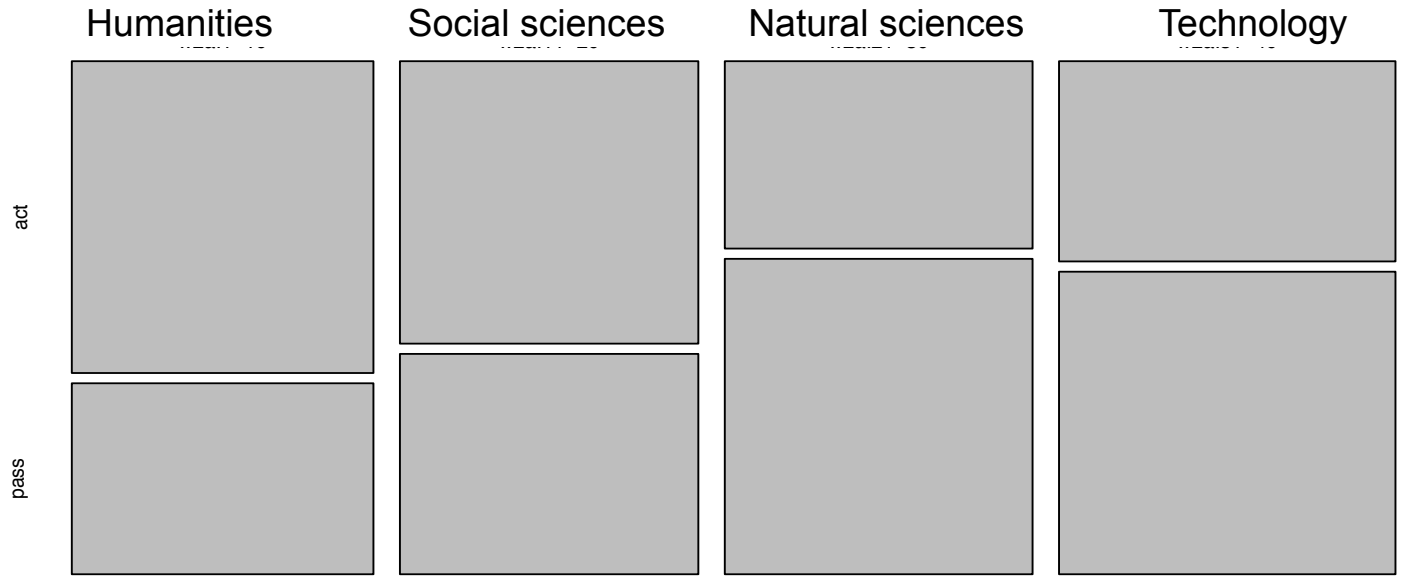


Figure 1: Actives and passives across 15 academic Englishes by subdiscipline

Previous research

Surprising discovery that regional variation is not really *that* important

BUT

- Focus on overall frequency of active vs. passive
- No information on *internal, linguistic* factors that predict the alternation

Seoane and Hundt (submitted) → qualitative analysis of factors in ENL varieties (BrE, AmE, CanE, IrE, AusE, NZE)

Research question: differences in authorial presence across subdisciplines?

→ Informalisation rather than differences in personalisation

Today: probabilistic modelling of internal factors treating passive-active as an alternation (i.e. a choice context)

Englishes selected

ENL: BrE, AmE

ESL: SingE, HKE, PhilE, IndE, FijE

	Phase 2	Phase 3	Phase 4
SingE			----------
HKE		----------	
PhilE		----------	
IndE		----------	
FijE	----------	----------	

Figure 2: Developmental phase of contact varieties of English (Schneider, 2007)

Substrate influence?

	Singapore	Hong Kong	India	Pilippines	Fiji
Periphrastic	-	-	+	-	-
Inflectional	-	-	-	-	+
Other	+	+	-	-	-

Figure 3: Passives across substrates in L2-Englishes

Factors

Complexity

they will have been being chased

Learners: tendency to avoid complexity, ESL varieties expected to simplify

(see Kortmann & Szmrecsanyi 2009 or Thomason 2013)

Animacy

Animate subjects preferred over inanimate ones (see animacy hierarchy, Silverstein 1976)

Regoinal variation in effect size of animacy (see e.g. Hinrichs & Szmrecsanyi 2008, Bresnan & Hay 2008 or Bresnan & Ford 2010; papers at this workshop)

Weight

Rearrange information so that heavy constituents are shifted out of subject position (Behagel 1909; papers at this workshop);

cognitive factor more important in ESL varieties?

Research questions

Will we see a divide into first- and second-language varieties when it comes to the internal factors predicting the choice of a passive over an active verb phrase (and if so, at what level)?

More specifically, will we find regional differences in the role that factors such as ‘animacy’ play as predictors?

For the second-language varieties, can we observe possible influence of substrate languages or the process of second-language acquisition in the ranking of factors that play a role in predicting voice alternation?

Data and methodology

Data and retrieval

- Academic writing section (published) of ICE

  10 x 2,000 words per sub-discipline and variety

  > approx. 80,000 words per variety

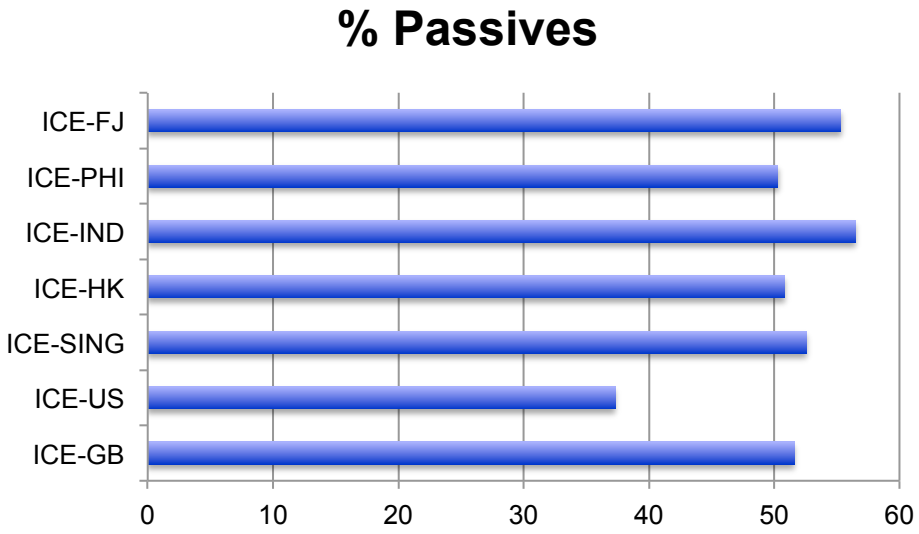
  > approx 560,000 overall

- Corpora are parsed (dependency parser, Schneider 2008)
  → automatic retrieval (see Hundt et al., 2016)

Automatically retrieved constructions

corpus	actives	passives
ICE-GB	1185	1265
ICE-US	1670	994
ICE-SIN	1133	1255
ICE-HK	1570	1620
ICE-IND	1117	1449
ICE-PHI	1374	1392
ICE-FJ	1118	1383

Automatically retrieved constructions



Analysis for internal factors: 100 randomly retrieved actives and passives each per variety; manually exclude false positives

Examples of false positives

Active transitives

① … more Fijian businesses would *bite* dust if Government did not act now (ICE-FJ W2A 011)

② Therefore it can be assumed that those who *enter* a university have chosen to do so … (ICE-SL W2A 004)

③ One of the problems I *wish to address* is the degree to which Frankish uncial in the late eighth and the ninth centuries is indeed artificial rather than natural. (ICE-GB W2A 008)

Data set of 1285 instances (610 actives, 675 passives)

Coding for factors

1285 relevant data sets

- Complexity of the VP

  simple (TENSE only) vs. complex (TAM markers)

- Semantics of the subject and object/ *by*-agent

  animacy (animate – inanimate – unclear)

  ① The Griffiths Report was not received with great enthusiasm by the Conservative government … (ICE-GB W2A 013)

  ② … and in the stress of war Rome conceded what they had sought. (ICE-GB W2A 001)

  ③ So that neither *the user or system* is overwhelmed by large result sets, the size of result sets is limited to 100 items … (ICE-US W2A 038)

Coding for factors

- Weight (length of subject and object / *by*-agent)

  number of words (0, 1, 2, 3, 4-9, 10-15, 16-20, >20)

  ① … [the high amplitude specular signal from the 0 probe]$_S$ emphasises [the surface of the defect]$_O$. (ICE-GB W2A 031)

Results

Complexity of the VP

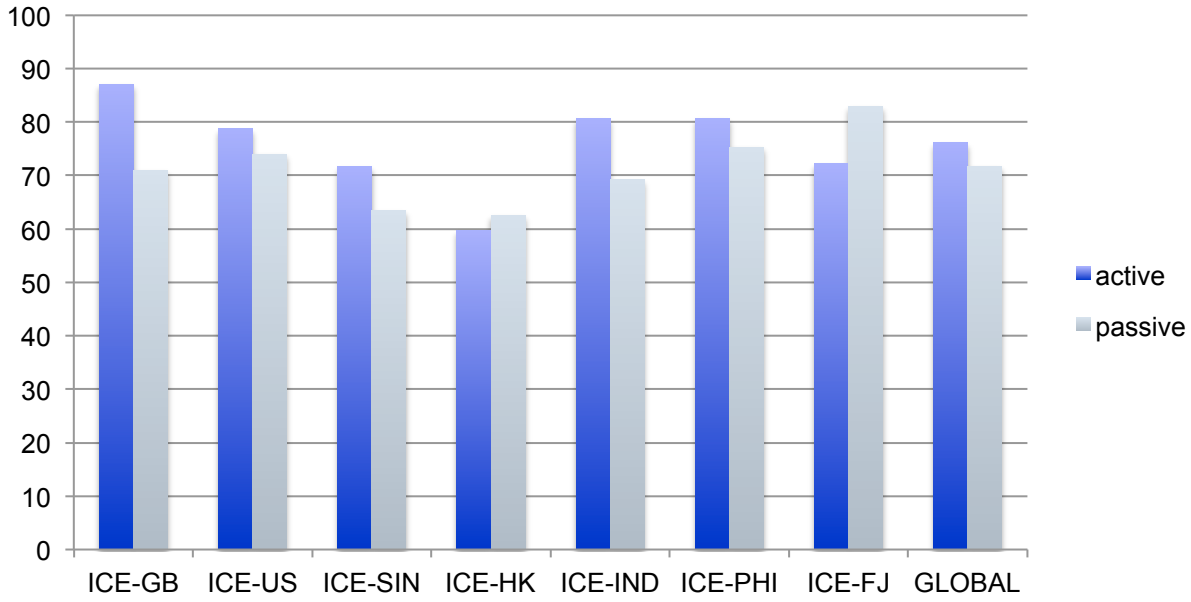


Figure 4: Percentage of simple VPs (present and past)

Semantics of the subject

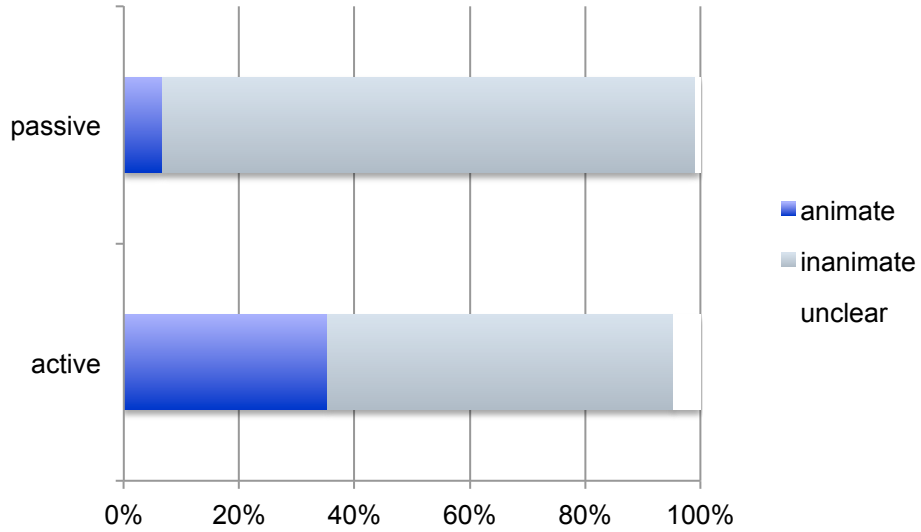


Figure 5: Relative frequency of animate vs. inanimate subjects (all varieties)

Semantics of the object / *by*-agent

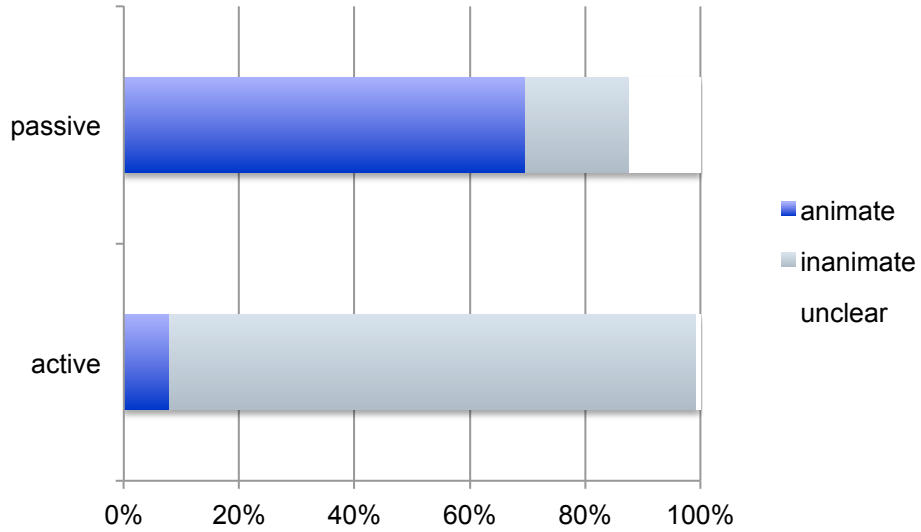


Figure 6: Relative frequency of animate vs. inanimate objects (all varieties)

Weight: Length of the subject

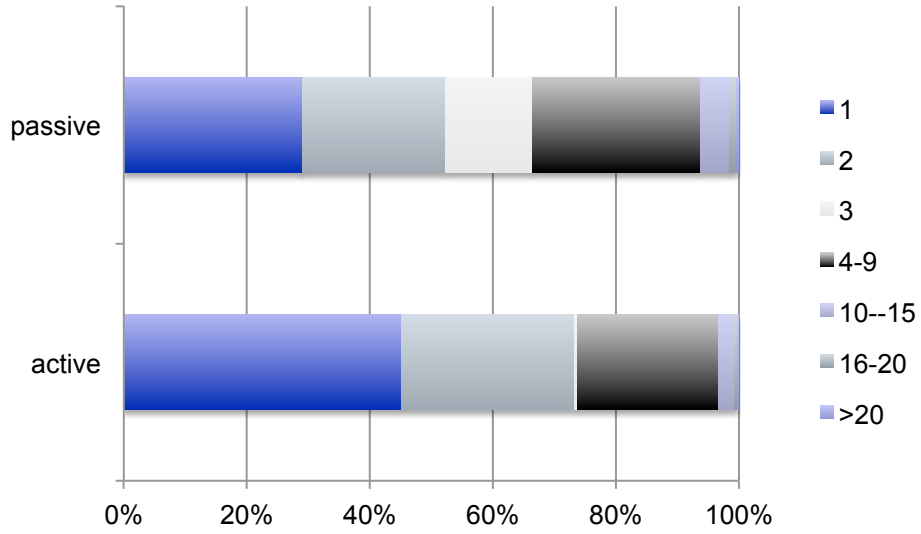


Figure 7: Length (words) of the subject (all varieties)

Weight: Length of the object / *by*-agent

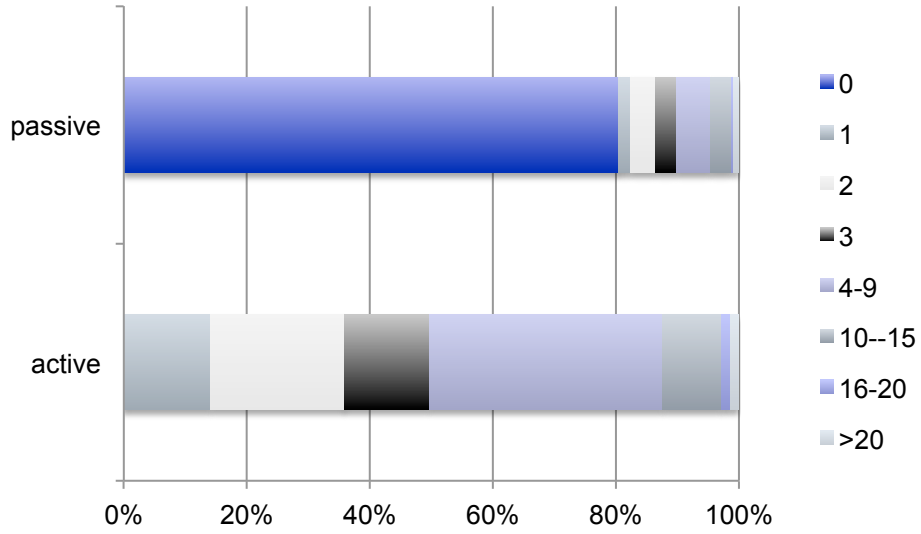


Figure 8: Length (words) of the object (all varieties)

ANOVA test of variance

Global variation

	F-value	*p*-value	η^2	Ranking
VP.structure	14.04	<.001*	0.03	3
SemanticsS	226.14	<.001*	0.14	2
SemanticsO	666.70	<.001*	0.40	1
Length.subject	*1.02*	*0.31*	*0.0003*	
Length.object	102.16	<.001*	0.03	3

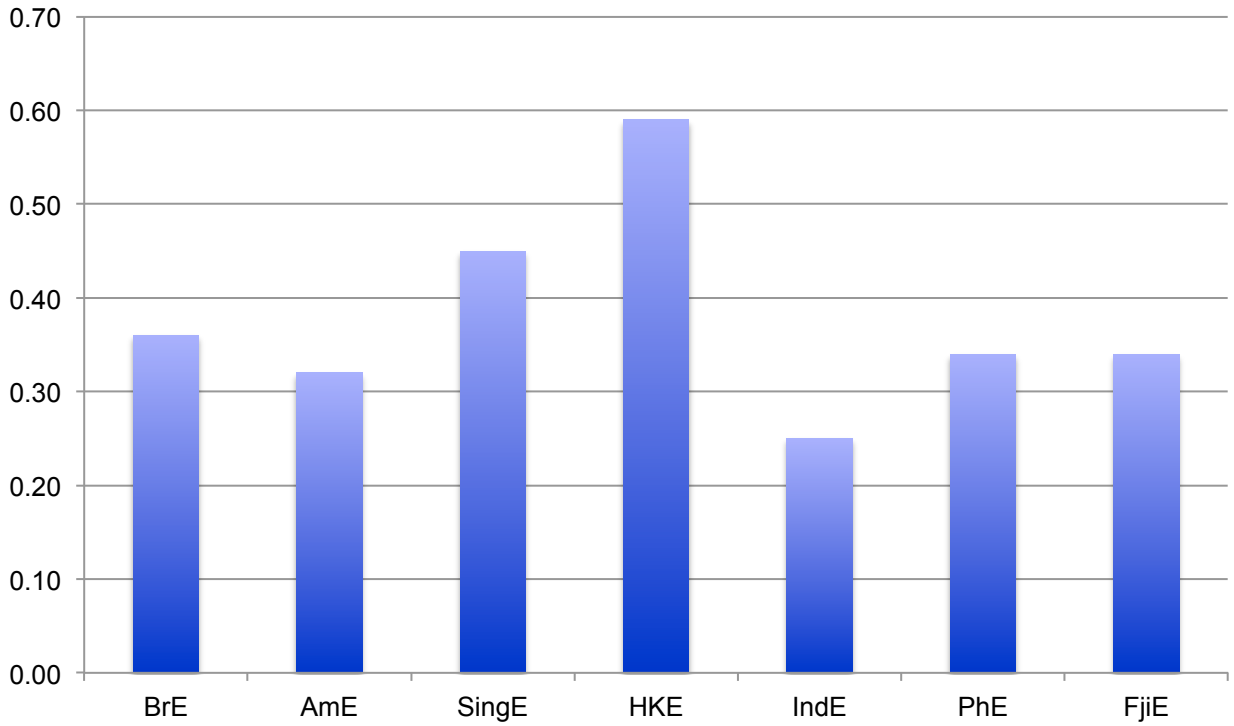
ANOVA test of variance

Inner Circle vs. Outer Circle?

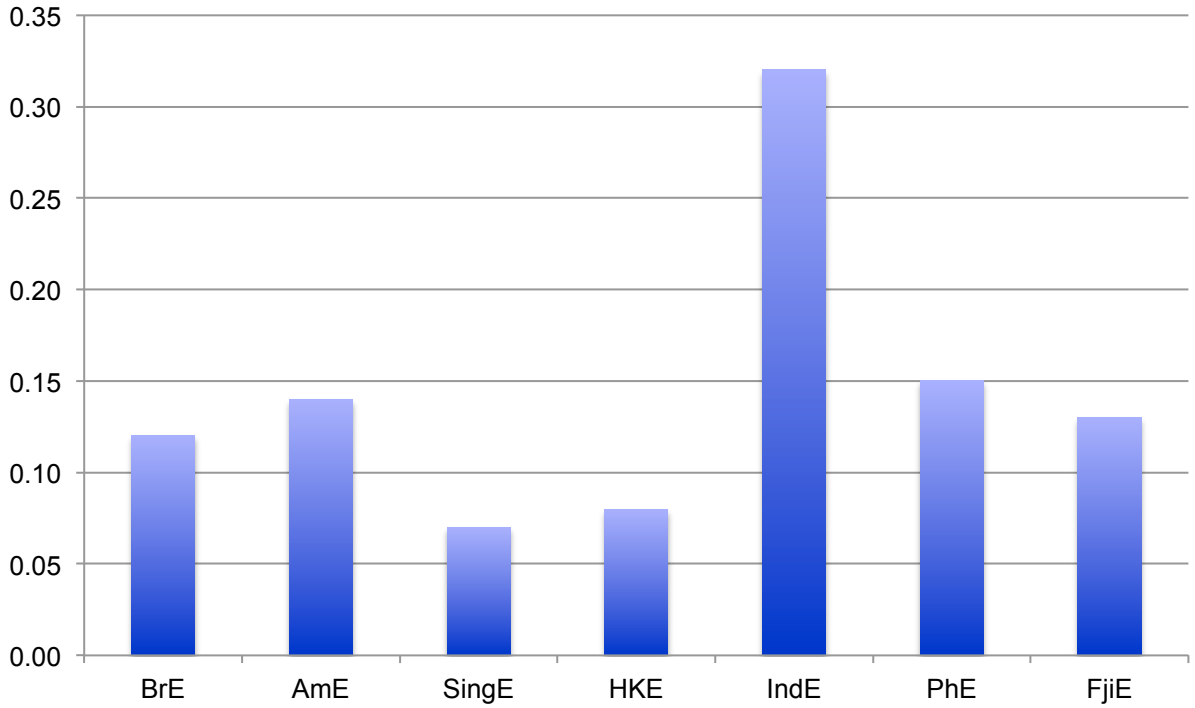
inner circle	*F*-value	*p*-value	η^2	Ranking
VP.structure	10.81	<.001*	0.07	3
SemanticsS	60.17	<.001*	0.14	2
SemanticsO	149.68	<.001*	0.34	1
Length.object	20.95	<.001*	0.02	4

outer cricle	*F*-value	*p*-value	η^2	Ranking
VP.structure	6.30	<.001*	0.02	4
SemanticsS	163.24	<.001*	0.13	2
SemanticsO	515.07	<.001*	0.42	1
Length.object	81.02	<.001*	0.03	3

Ranking of effect size: Semantics of object



Ranking of effect size: Semantics of subject



Order of constraints across ESL varieties

	ICE-SIN	ICE-HK	ICE-IND	ICE-PHI	ICE-FJ
VP.structure	3	3	3	3	3
SemanticsS	2	2	1	2	2
SemanticsO	1	1	2	1	1
Length.subject		**4**			
Length.object	3	5	3	4	4

Discussion

Research questions

Will we see a divide into first- and second-language varieties when it comes to the internal factors predicting the choice of a passive over an active verb phrase (and if so, at what level)?

→ Surprising homogeneity

→ No difference between BrE and AmE (but: stable if more ENL varieties added into the picture?)

→ (Slight) differences in the effect size (notably: IndE)

→ Occasional difference in the ranking of factors (VP complexity more important for ESL varieties than ENL – only globally, though, not necessarily for individual ENL varieties)

Research questions

For the second-language varieties, can we observe possible influence of substrate languages or the process of second-language acquisition in the ranking of factors that play a role in predicting voice alternation?

→ IndE expected to be closest to BrE and AmE, but in fact v different

Conclusion

Overarching research questions of the workshop

- What is the extent to which varieties of English share a probabilistic grammar that predicts variation patterns across different varieties?
- How do we evaluate overall probabilistic similarity between varieties?

Probabilistic indigenization?

The typical pattern to emerge from such studies is that of a core probabilistic grammar with variety-specific peculiarities at a more fine-grained level of analysis.

→ Similarities appear to be more marked than differences

→ These can be VERY subtle

→ They don't easily align with Schneider's (2007) model – SingE expected to diverge MORE from BrE and AmE than it does

→ More background needed to account for e.g. diverging behaviour of IndE

Probabilistic grammar

What does it all *mean*?

Levels of description/modelling and how they connect.

An analogy from biology…

World Englishes and statistical modelling

Visual/surface similarity vs. morphological/ evolutionary similarity

(a)
Hyrax – guinea pig
(b)
Hyrax – elephant, aardvark (Afrotheria)

References

Behagel, Otto. 1930. Von deutscher Wortstellung. *Zeitschrift für Deutschkunde* 44: 81-89.

Biber, Douglas and Edward Finegan. 1989. Drift and evolution of English style: a history of three genres. *Language* 65: 487-517.

Biewer, Carolin. 2009. Passive constructions in Fiji English: A corpus-based study. In Andreas H. Jucker, Daniel Schreier and Marianne Hundt, eds. *Corpora: Pragmatics and Discourse*. Amsterdam: Rodopi, 361-377.

Biewer, Carolin. 2015. *South Pacific Englishes. A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English*. Amsterdam: Benjamins.

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: 245–59.

Bresnan, Joan and Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English." *Language* 86,1: 168–213.

Davies, Mark and Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36(1): 1-28.

Deterding, David, Jenny Wong and Andy Kirkpatrick. 2008. The pronunciation of Hong Kong English. *English World-Wide* 29(2): 148-175.

Deterding, David. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.

Geraghty, Paul, France Mugler and Jan Tent, eds. 2006. *Macquarie Dictionary of English for the Fiji Islands*. Macquarie University: The Macquarie Library.

Geraghty, Paul. 2008. *Fijian*. Victoria: Lonely Planet Publications.

References

Gries, Stefan Th. and Tobias Bernaisch. 2016. Exploring epicentres empirically. Focus on South Asian Englishes. *English World-Wide* 37(1): 1-25.

Hinrichs, Lars and Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11: 437-74.

Hundt, Marianne, Gerold Schneider and Elena Seoane. 2016. The use of the *be*-passive in academic Englishes: Local vs. global usage in an international language. *Corpora* 11(1): 31-63.

Hundt, Marianne and Christian Mair. 1999. Agile and uptight genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2): 221-242.

Hundt, Marianne, Lena Zipp and André Huber. 2015. Attitudes towards varieties of English in Fiji: A shift to endonormativity? *World Englishes* 34(3): 688-707.

Hundt, Marianne. 2004. The passival and the progressive passive: a case study in layering in the English aspect and voice systems. In Hans Lindquist and Christian Mair, eds. *Corpus Approaches to Grammaticalization in English*. Amsterdam: Benjamins, 79-120.

Hundt, Marianne. 2007. *English Mediopassive Constructions*. Amsterdam: Rodopi.

Hundt, Marianne. 2013. Using web-based data for the study of global English. In Manfred Krug and Julia Schlüter, eds. *Research Methods in Language Variation and Change*. Cambridge: University Press, 158-177.

Kachru, Yamuna. 2006. *Hindi*. Amsterdam: Benjamins.

Keenan, Edward L. 1985. Passive in the world's languages. In Timothy Shopen, ed. *Language Typology and Syntactic Description*, Volume 1. Cambridge: Cambridge University Press, 243-81.

Kondo, Takako. 2005. Overpassivization in second language acquisition. *International Review of Applied Linguistics in Language Teaching (IRAL)* 43: 129–161.

References

Kortmann, Bernd and Benedikt Szmrecsanyi. 2009. World Englishes between simplification and complexification. In Thomas Hoffmann and Lucia Siebers, eds. *World Englishes. Problems, Properties and Prospects*. Amsterdam: Benjamins, 263-286.

Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study.* Cambridge: Cambridge University Press.

Lynch, John. 1998. *Pacific Languages: An Introduction*. Honolulu: University of Hawai'I Press.

Matthews, Stephen and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London and New York: Routledge.

McFarland, Curtis D. 2008. Linguistic diversity and English in the Philippines. In Ma. Lourdes S. Bautista and Kingsley Bolton, eds. *Philippine English: Language and Literary Perspectives*. Hong Kong: Hong Kong University Press, 131-156.

Sandahl, Stella. 2000. *A Hindi Reference Grammar*. Leuven: Peeters.

Schachter, Paul and Fe. T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World.* Cambridge: Cambridge University Press.

Schneider, Gerold. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. PhD, University of Zurich.

Seoane, Elena and Lucía Loureiro-Porto. 2005. On the colloquialization of scientific British and American English. *ESP Across Cultures* 2: 106-118.

Seoane, Elena and Marianne Hundt (submitted). Styling academic Englishes: voice alternation and authorial presence across disciplines.

Seoane, Elena. 2006. Changing styles: on the recent evolution of scientific British and American English. In Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendl, eds. *Syntax, Style and Grammatical Norms: English from 1500-2000.* Bern: Peter Lang, 191-211.

References

Shibatani, Masayoshi. 1988. Voice in Philippine languages. In Masayoshi Shibatani, ed. *Passive and Voice*. Amsterdam: Benjamins, 85-142.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon, ed. *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies, 112-171.

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller and Melanie Röthlisberger. Forthcoming 2016. Around the world in three alternations: modelling syntactic variation in varieties of English. *English World-Wide* 37(2).

Xiao, Richard, Tony McEnery and Yufang Qian. 2006. Passive constructions in English and Chinese. A corpus-based contrastive study. *Languages in Contrast* 6(1): 109-149.

Zipp, Lena. 2014. *Educated Fiji English. Lexico-grammar and Variety Status*. Amsterdam: Benjamins.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Rosenbach, Anette. 2007. Animacy and grammatical variation: findings from English genitive variation. *Lingua* 118: 151-71.

Seoane, Elena. 2009. Syntactic complexity, discourse status and animacy as determinants of grammatical variation in English". *English Language and Linguistics* 13(3): 365-384.

Thomason, Sarah G. 2013. Innovation and contact: the role of adults (and children). In Daniel Schreier and Marianne Hundt, eds. English as a Contact Language. Cambridge: University Press, 283-297.

Wasow, Thomas and Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In Gunter Rohdenburg and Britta Mondorf, eds. *Determinants of Grammatical Variation in English*. Berlin: de Gruyter, 119-154.

University of Zurich[UZH]

English Department URPP Language and Space Universidade Vigo

Acknowledgements

Swiss National Science Foundation (SNSF) grant (IZK0Z1_149005) that enabled joint work on the project in August 2013.

Spanish Ministry of Science and Innovation & Autonomous Government of Galicia (Grant Nos., FFI2011-26693-C02-01 and FFI2011-26693-C02-02; CN2012/012).

Gerold Schneider for retrieving the raw data.

Simone Pfenninger for help with the statistics