

# **Academic Literacies and EAP Learner Corpus Research**

**Lynne Flowerdew**

**Birkbeck, University of London**

# Aim of Presentation

- To review learner corpus studies and their pedagogic applications/implications for disciplinary writing over past five years
- To see where these studies interface (or not) with the Text Histories approach to disciplinary writing

**Worlds apart or overlapping spheres?**

- End with: **Motivation for presentation**

# Text Histories Approach

- Described as:  
**longitudinal**                      **text-oriented**                      **ethnographic**
- Aims to capture:  
**Textual** and **contextual** data
- Used for describing and analysing:

**Process of peer review for scholarly publication**  
**Collaborative writing (with a view to publication)**  
**'networked activity'**

# Learner Corpus Studies

**What are the different orientations?**

# Text-oriented (Ia)

- Callies (2013) research on interlanguage of L1 German advanced students in area of 'agentivity' (construal of author's own work and that of other researchers)
- Genre and discipline: research reports in linguistics
- Corpora used:
  - Subset of CALE (Corpus of Academic Learner English)
  - Subset of English native-speaker writing from MICUSP
    - comparable across academic level, discipline and text-type
  - size of corpora: 62,000 words each (mainly qualitative analysis)

# Text-oriented (Ib)

## Findings

- Overrepresentation of 1<sup>st</sup> pers. pronouns
- Overrepresentation of subject placeholders with passivised reporting verbs, e.g. *It can be observed..*
- Inanimate subjects found to be underrepresented, e.g. *The importance of hedging has been demonstrated...*
- Conclusion: L2 writers lack familiarity with disciplinary conventions (not due to greater involvement or higher degree of subjectivity)

# Ethnographically-oriented (Ia)

Sing's (2016) content analysis of advanced L1 German advanced business students' use of technicality markers

- 'Technicalising': two-stage process, consisting of naming a term and subsequently embedding it in taxonomic relationships
  - Discourse functions: defining; exemplifying; explaining
- Corpus used:
  - One-million-word corpus of advanced business English (ABE)
  - Divided into four sub-corpora
  - 400 seminar papers produced by advanced business students
  - No native-speaker control corpus used
- ABE corpus: **localised** collection of texts

# Ethnographically-oriented (Ib)

Sing emphasises use of contextual information for analysis of data (compiler-cum-analyst role familiar with ethnographic context )

- Students rely heavily on structure provided in source material
- Level of categorisation rather than language that students are in need of assistance with
- Student writers seem preoccupied with taxonomies rather than developing ideas (low lexical density)  
e.g. Joint Ventures (=section heading) A joint venture **can be defined as** a company which is owned corporately by two or more parties [ABE\_Business]

# Quasi longitudinally-oriented (1a)

Miller and Pessoa (2018) examined six rhetorical functions (elaboration, reasoning) across different levels of disciplinary writing

- Genre and discipline: information system project reports across **junior and senior project reports**
- Corpora used:
  - Frown corpus containing a wide variety of genres and topics
    - Used as a reference corpus to identify rhetorical functions typifying those in their corpus of **expert texts**
  - **junior and senior project reports; collaborative**
  - Size of corpora: 17/18 texts with a mean length = 400 words 9

# Quasi longitudinally-oriented (Ib)

## Findings

- Students were able to incorporate in their own writing most of the functions found in the model samples without any explicit instruction
- Senior students showed greater client awareness
- Areas of difficulty: reporting about processes team undertakes in developing project focused entirely on past steps without describing how team planned to address problems

# Quasi longitudinally-oriented (II)

Hafner and Wang (2018) investigated boosters in a corpus of legal academic writing in English by Hong Kong undergraduate law students **across four years of study**

- Genres: case notes, judgment, essay, problem-question
- Data collection:
  - total: 1,037,387 from genres in 1<sup>st</sup> – 4<sup>th</sup> years of study
  - Collected highly-graded assignments
  - **Consulted their lawyer co-investigators** to draw the line between distinction and merit assignments

## Findings

- Students used fewer boosters as they received more years of training – adapting their writing to disciplinary values

# Pedagogically-oriented

## Pedagogic applications: online tools

- Birch-Becaas and Cooke (2012) describe how a small corpus of French novice writers' drafts of RA introductions used
  - revised versions included
  - advantages of corpora and tools for doctoral students who have no access to expert writers and no funding for text-editing services
- Nordrum and Eriksson's (2015) online tool for data commentary in science writing
  - Incorporates texts of different levels and quality

## Summing up: learner corpus studies and applications

- Much less emphasis on NS vs. NNS writing
- More emphasis on expert vs L2 writing (or EAL writing)
- More attention paid to what students can do rather than a focus on a deficit model
- Value of corpus findings
  - Recurrent lexico-grammatical patterns indicate where difficulties lie – useful for ESP needs analysis
- Ethnographic dimension can provide insights for data collection and analysis and interpretation of data
- Quasi-longitudinal dimension indicates students' progress across years of study

# Is there any interface between learner corpus studies and Text Histories approach?

- Learner corpus studies
  - More contextually-oriented
  - More (quasi)-longitudinal

## However

- Focus on writing as product
- Difficult to capture:
  - Writing as process
  - Collaborative nature of writing as a networked activity

# Motivation for presentation

# PhD study on collaboration practices of supervisors and their doctoral students

**Pascal Matzler:**

Paper of a case study presented at Faces of English Conference in Hong Kong 2017

**'Mentoring by co-authoring in the natural sciences:  
a text history'**

Aim of study:

To use **Text Histories** to trace how one doctoral student and their supervisor work together to author a research article

# Data collection for PhD (Matzler 2017)

Hello Chet  
New version!  
Ciao  
Tsai

Hello Tsai,  
new version attached. we are getting there. definitely. we still need a chat about a couple of points. but we are definitely getting there.  
meeting: what about Thursday 10am? i have plenty of visitors until wednesday and on wednesday there's Roberto's presentation. busy week. let me know  
Chet

381 advective (equation (4) and (8)). In fact, as discussed by Schuttelaars and  
382 De Swart (2001), who compared the advectively dominant and the diffusively  
383 dominant sediment transport in a idealized tidal basin, the same equilibrium  
384 state is derived for both sediment transport modes. Certainly, the difference  
385 is that the advective sediment transport component is zero where the veloc-  
386 ity is zero (e.g. the closed upstream boundary), while the diffusive transport  
387 is always present unless the spatial gradient of the tide-averaged sediment  
388 fluxes is zero. However, the "diffusive effect" driven by the back and forth

381 advective (equation (4) and (8)). In fact, as discussed by Schuttelaars and  
382 De Swart (2001), who compared <sup>effect of advective and diffusive sediment transport</sup> ~~the advectively dominant and the diffusively~~  
383 ~~dominant sediment transport~~ <sup>an</sup> in a idealized tidal basin, the same equilibrium  
384 state <sup>is</sup> derived for both sediment transport modes. Certainly, ~~the difference~~  
385 ~~is that~~ the advective sediment transport component is zero where the veloc-  
386 ity is zero (e.g. the closed upstream boundary), while the diffusive transport  
387 is ~~always~~ present unless the spatial gradient of the tide-averaged sediment  
388 fluxes is zero. However, the "diffusive effect" driven by the back and forth

Note - [.. 12/03/2017, 4:03:45 p.m. x  
i think you have to clarify and define what you mean with 'advective and diffusive'

# Types of data collected

RA drafts

Email dialogue

On-script notes

Insertions, deletions and substitutions

381 advective (equation (4) and (8)). In fact, as discussed by Schuttelaars and  
382 De Swart (2001), who compared <sup>effect of advective and diffusive sediment transport</sup> ~~the advectively dominant and the diffusively~~  
383 ~~dominant sediment transport~~ <sup>an</sup> in a idealized tidal basin, the same equilibrium  
384 sta<sup>s</sup> derived for both sediment transport modes. Certainly, ~~the difference~~  
385 ~~is that~~ the advective sediment transport compo<sup>s</sup>  
386 ity is zero (e.g. the closed upstream boundary),  
387 is ~~always~~ present unless the spatial gradient of  
388 fluxes is zero. However, the "diffusive effect" di

Note - [.. 12/03/2017, 4:03:45 p.m. x  
i think you have to clarify and  
define what you mean with  
'advective and diffusive'

Dialogue/interaction in meetings

16.03.2017 at 07min50sec

"This brings us to the key question of today: the definition of advective and diffusive.

Yes.

We need to do something about that, because it comes up so late, and it's never properly defined.

# Data collection for PhD (Matzler 2017)

- Significance of “post-it” notes
  - Supervisor makes more profound reflections on more general problems
  - Supervisor uses notes to structure meeting with student and guide discussion
- Advantages of analysing various forms of data in parallel
  - Drafts
  - Written feedback
  - Meeting transcripts

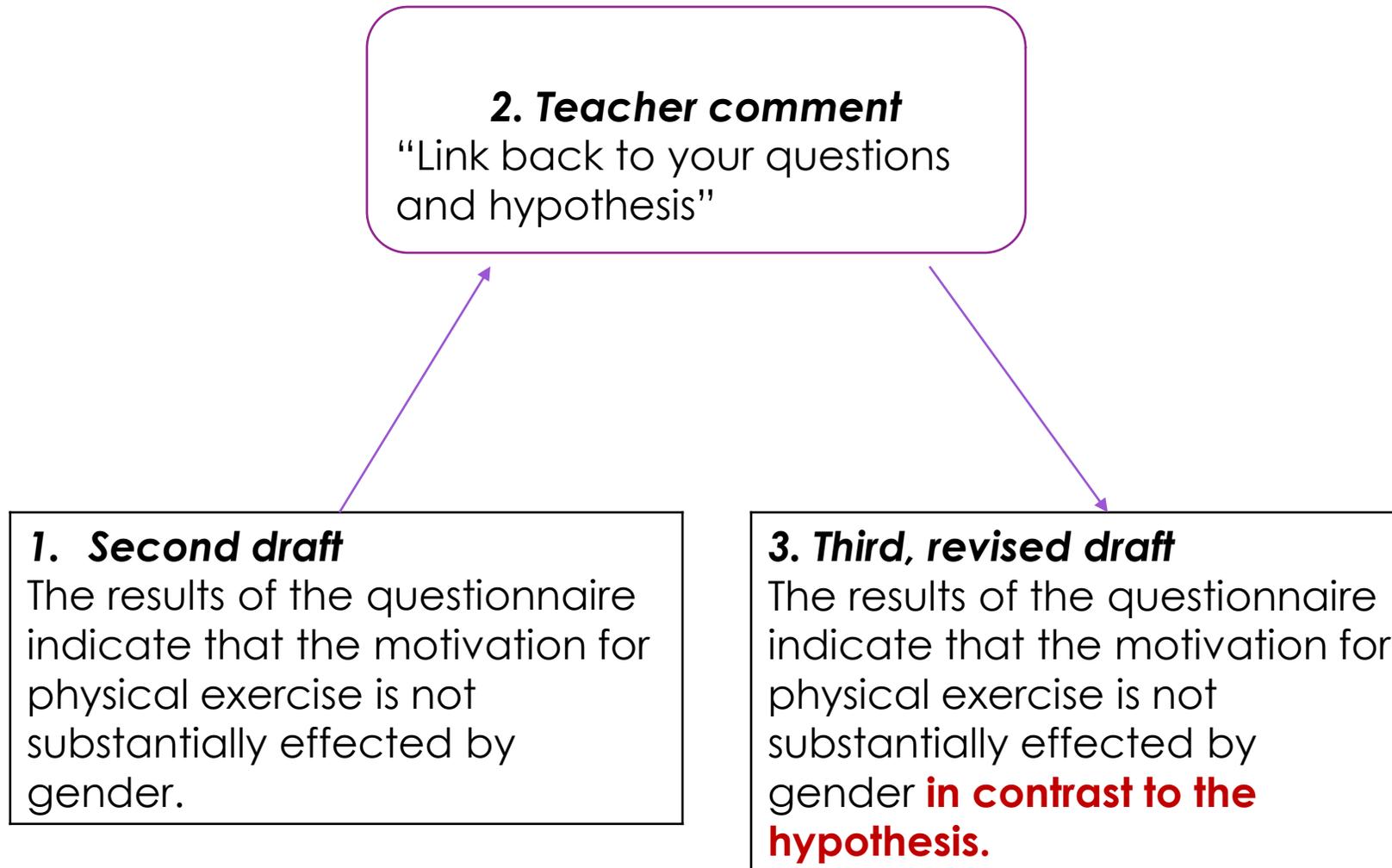
# Overlapping worlds?

# Learner corpora and writing as process

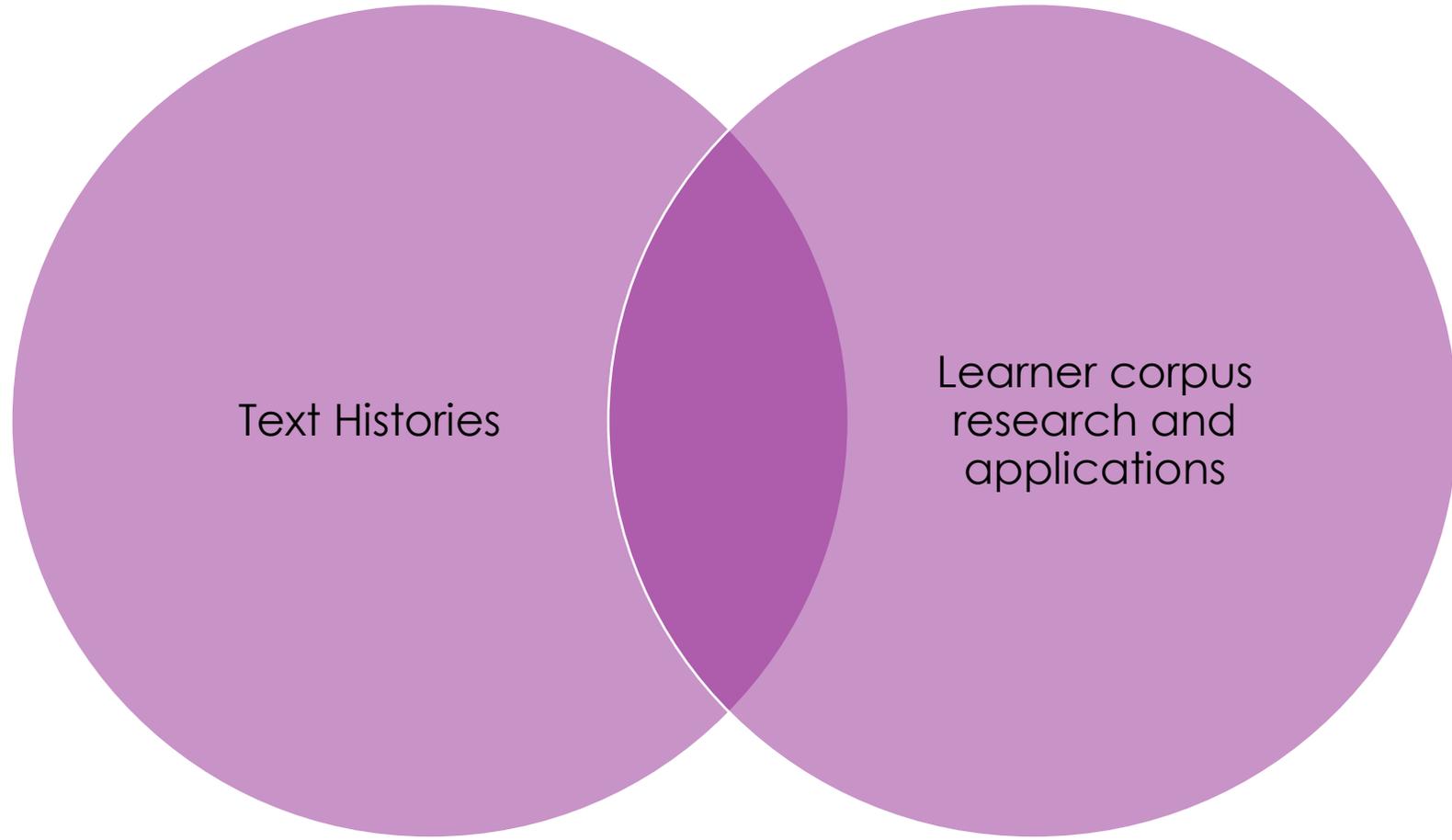
Eriksson et al. (2016) on MUCH (Malmö University-Chalmers **Corpus of Academic Writing as Process**)

- Corpus designed to capture ‘**writing as process**’
- Includes several drafts, student self-reflective papers, and teacher and peer feedback
- Uses **pan-European CLARIN infrastructure**
- Tagging for rhetorical structures and comments of different types – **advantage over case study approach**

# Learner corpora and writing as process



# Contextual, Longitudinal, Writing as Process



## Reference:

Flowerdew, L. (2019). The Academic Literacies Approach to Scholarly Writing: a view through the lens of the ESP/Genre approach. *Studies in Higher Education*.

**Thank you!**