

Learner Language, Discourse and Interaction

Tony McEnery (with Issy Clarke and Gavin Brookes)

 @tonymcenery.bsky.social

- Our data
- Our focus
- Our problem
- Our solution
- Our findings

STUDIES IN ENGLISH LANGUAGE

The aim of this series is to provide a framework for original studies of English, both present-day and past. All books are based securely on empirical research, and represent theoretical and descriptive contributions to our knowledge of national and international varieties of English, both written and spoken. The series covers a broad range of topics and approaches, including syntax, phonology, grammar, vocabulary, discourse, pragmatics and sociolinguistics, and is aimed at an international readership.

Learner Language, Discourse and Interaction

'This is a monumental study, situated at the intersection of second language acquisition and discourse analysis. The authors spared no effort in developing and applying corpus linguistic methods – both quantitative and qualitative – that shed light on the sources of variation within speech produced by native and non-native speakers of English. The findings are a treasure trove of important contributions to the cumulative body of knowledge within applied linguistics and beyond.'

Jesse Egbert, Associate Professor of Applied Linguistics at Northern Arizona University

'This book provides an insightful exploration of spoken learner discourse through several novel corpus-based approaches. I heartily recommend it to anyone interested in advancing research on corpora, learner language, and discourse.'

Eric Prigal, Professor and Head, Department of English and Communication, The Hong Kong Polytechnic University

'This book offers a ground-breaking exploration of under-researched areas in learner language interaction such as discourse units and turns using short-text Multi Dimensional Analysis (MDA). The study of the L1 interlocutor and the consideration given to the emerging functions at the micro and macro levels of analysis represent a step forward in research methodology in this field.'

Pascual Pérez-Paredes, Professor in Applied Linguistics and English Linguistics, Universidad de Murcia



This title is also available as Open Access on Cambridge Core at www.cambridge.org/core

Cover illustration: Decorative stone wall.
Roberto Machado Noa/Getty Images.



ISBN 978-1-009-20895-6



9 781009 208956 >

McEnery, Clarke
and Brookes

Tony McEnery, Isobelle Clarke
and Gavin Brookes

Learner Language, Discourse and Interaction

A Corpus-Based Analysis of Spoken English

Learner Language, Discourse
and Interaction

CAMBRIDGE

STUDIES IN ENGLISH LANGUAGE

Three Corpora used in this Talk

- Trinity Lancaster Corpus - 1,737,822 word annotated subset
- Trinity Lancaster L1 Corpus - 901,085 word annotated subset
- British National Corpus 2014 (spoken) – 3,265,194 word annotated subset
- All coded for discourse units
- Main focus of investigation – task, grade of exam and proficiency

L2 users: English proficiency

Three levels

- Advanced: Grades 10, 11, 12 (CEFR: C1 & C2)
- Intermediate: Grades 7 & 8 (CEFR: B2.1 & B2.2)
- Threshold: Grade 6 (CEFR: B1.2)

How measured?

- ✓ Proficiency by task (A-D)
- ✓ Proficiency overall (A-D)

Four tasks - overview

	Topic familiarity	Interlocutor roles	Interactiveness	Formality
PRES	pre-selected topic	candidate-led	monologic	(semi-)formal
DISC	pre-selected topic	jointly-led	dialogic	semi-formal
INT	general topic	candidate-led	dialogic	semi-formal
CONV	general topic	jointly-led	dialogic	semi-formal

- Interactions dialogic
- Emphasis on continuing the conversation

Discourse

- Meaning above the level of the turn
- Macrostructure v. microstructure
- Differing views on what the macrostructures are
- Our approach – broadly functionalist (after Schiffrin, 1994), with form relating to function. However, also inventory like – but our inventory of features is more atomic.
- What are our macrostructures?
- We will call these discourse units.

Discourse Units

- Atomic discourse macrostructures which are functionally meaningful.
- Collaborating with Doug Biber and Jesse Egbert at NAU, work supported by TCL
- Initial Data – BNC 2014 spoken (approximately 5,000,000 words annotated)
- Useful as it establishes a benchmark for everyday conversational competence

Discourse Units

- Coded by a team of coders, mainly in the UK, though some at NAU
- Coders trained to produce plausible analyses
- Two primary goals: (1) attempt to identify recognizable shifts in communicative function to identify speech event boundaries, and (2) attempt to determine the primary communicative function for each speech event
- We initially focus on 1 in this talk as part of a bottom up approach to discourse function and then shift to consider 2 as a top down approach later
- Rule of thumb – discourse units tend to be around ten turns long.

Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021) 'Identifying and describing functional discourse units in the BNC Spoken 2014', *Text and Talk*, 41 (5-6), pp 715-737.

Problems

- How do we approach the uppermost level of discourse micro-structure – the turn?
- But how can we develop a functional view of the turn which is comparable to that of the discourse unit?
- How could this view then show how co-construction of discourse works?

A possible solution – the multi-dimensional approach

- In principle, this explicitly meshes micro and macro levels – in the Biber paradigm the configuration of micro level features (lexis/part of speech) are strongly indicative of macro level features (e.g. register).
- Could we not take the same approach to discourse units and turns?
- If so, we would abandon the labelling of Egbert et al and, instead, simply look at the discourse unit boundaries and try to make sense of what was in them.
- We would do the same with the turns in the data, separately.

What is Multi-Dimensional Analysis?

A methodological approach, resting on frequency of co-occurrence of linguistic features, used to:

1. identify the most common patterns of linguistic variation across a corpus of texts
2. identify the major communicative functions across a corpus of texts representing a particular linguistic variety
3. compare corpora according to multiple patterns of linguistic variation and communicative functions (often focused on register variation)

The problem of frequency

- Not as the technique stands – the problem of frequency in short texts.

E: <unclear text="you"/>

S: I'm fine ma'am

E: good

S: how are you ma'am?

E: I'm good thank you very much okay so erm good morning my name is <anon type="name"/> what's your name?

S: my name is <anon type="name"/>

E: <anon type="name"/>

S: yes ma'am

E: and you're a grade seven

S: yes

E: and can you just turn your oh you don't need to take it off

S: mm

E: okay thanks thanks for showing me that that's your ID card thanks <anon type="name"/>

- 68 words long
- BE as a main verb occurs 6 times
- Normalised frequency per 1,000 words of 88.2
- Frequency per 1,000 words of BE as a main verb is 37.64

Realising a multi-dimensional approach to sparse data

- The relative frequencies of features in short texts are not meaningful:
 - The relative frequencies of most grammatical forms become accurate in text samples of 500 (Passonneau et al., 2014) or 1000 words (Biber, 1993).
- Additionally, short texts tend to have few features, meaning that there are many zeros in the dataset, i.e. zero-inflation. This can mean that the data doesn't readily fit standard distributions.

A solution

- How could this be done? Replace the ‘engine’ of the mutli-dimensional approach, factor analysis, which runs on the fuel of normalised frequency, with one that runs on binary fuel!
- Multiple correspondence analysis (Benzécri, 1979; Clarke and Grieve, 2017).

Benzécri, J.P. (1979). Sur le calcul des taux d’inertie dans l’analyse d’un questionnaire. *Cahiers de l’Analyse des Données*, 4, 377–378.

Clarke, I. and Grieve, J. (2017) Dimensions of abusive language on Twitter. In *Proceedings of the First Workshop on Abusive Language Online* 1–10. Vancouver: ACL. <https://doi.org/10.18653/v1/w17-3001>

Dominance



+N

-N

Function A

		DU 1	Dimension 1	
--	--	------	-------------	--

Function B

Function C

DU 1			Dimension 2	
------	--	--	-------------	--

Function D

Function E

			Dimension 3	DU 1
--	--	--	-------------	------

Function F

Function G

		DU 1	Dimension 4	
--	--	------	-------------	--

Function H

Function I

			Dimension 5	DU 1
--	--	--	-------------	------

Function J



Interpretation

- Remember, MDA/MCA results show overlapping patterns of usage – look at extreme values to get a sense of a function
- Outliers on each dimension explored – first fifty examples used to propose a function. Second fifty examples then analysed to check whether the function holds across them.
- Two analysts doing the checking.
- The technique worked well – our initial functional analyses defied falsification through this procedure.
- A key advantage of the approach – we can use the same approach to analyse function at the discourse unit and the turn level separately. We can explore both micro and macro structure of discourse.

Looking at results

- I will narrow my focus to grade 6, 7 and 8 (CEFR B1/B2), focusing on the conversation (unplanned) and discussion (planned) tasks
- Three sets of results are presented – examiner turn, learner turn and discourse unit functions
- We will see how function interacts with a number of supplementary variables
- Let's begin by looking at the examiner turn functions and then focus in on dimension two

Dimension	Label	Features present
Dim 1 +ve	Long turns	Prediction modal, Infinitive, Private verb, Nominalisation, Quantifier, Auxiliary DO, Predicative Adjective, Subordinator, Definite article, General Determiner, Past tense, Analytic negation, Indefinite article, Contracted forms, Pronoun it, Third person singular verb, Coordinating Conjunction, General adverb, Preposition, Stative forms, First person pronoun, WH-word, Question mark, Subject pronoun, General verb, Second person pronoun, Attributive adjective, General noun.
Dim 1 -ve	Short turns	Only absent features (Subject pronoun, General noun, General verb)
Dim 2 +ve	Descriptive	Predicative adjective, Pronoun it, Contracted forms, Analytic negation, Stative forms, Demonstratives, Amplifier, Third person singular verb forms
Dim 2 -ve	Information seeking	WH-word, Auxiliary DO, Question marks, Second person pronoun
Dim 3 +ve	Guide future action	Nominalisation, Modal of prediction, Public verb, Quantifier, First person pronoun, Infinitive, General adverb, Definite article, Preposition
Dim 3 -ve	Discovering stance	WH-word, Auxiliary DO, Third person singular verb, Question, Predicative adjective, Past tense verb, Pronoun it, stative form

Dimension	Label	Features present
Dim 4 +ve	Stating stance	Analytic negation, Auxiliary DO, Negative interjection, Private verb, Third person pronoun, Infinitive, Subject pronoun, General verb
Dim 4 -ve	Discussing the here and now	Third person singular verb, Demonstrative, General determiner, Proper noun, Stative, Definite article, Preposition, Attributive adjective, General noun
Dim 5 +ve	Interjection (positive)	Positive interjection
Dim 5 -ve	Interjection (other)	Negative interjection, General interjection
Dim 6 +ve	Past orientation	Negative interjection, Indefinite article, Third person singular verb, Analytic negation, Third person pronoun, Past tense verb, Quantifier, Definite article, Attributive adjective, Amplifier, Preposition, General Noun
Dim 6 -ve	Future orientation	Modal of prediction, Contracted form, Infinitive, Predicative adjective, Demonstrative, General determiner, Second person pronoun, Subordinator, Positive interjection
Dim 7 +ve	Narrative	Third person pronoun, Negative interjection, Past tense verb, Public verb, Third person singular verb, General determiner, Subordinator
Dim 7 -ve	Stance seeking	Amplifier, Attributive adjective, Quantifier, Auxiliary DO, Private verb, Indefinite article, Second person pronoun

Dimension 2

Positive: Descriptive

so I'm not sure I mean I suppose if she's very academic maybe it is a better school for her

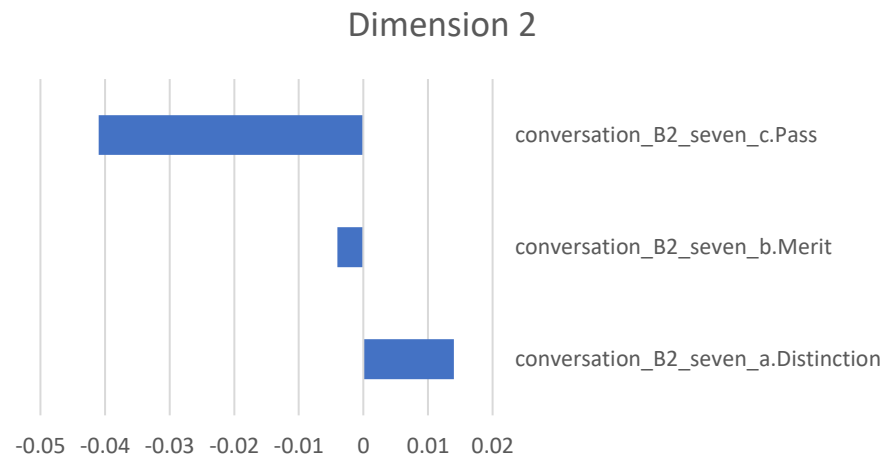
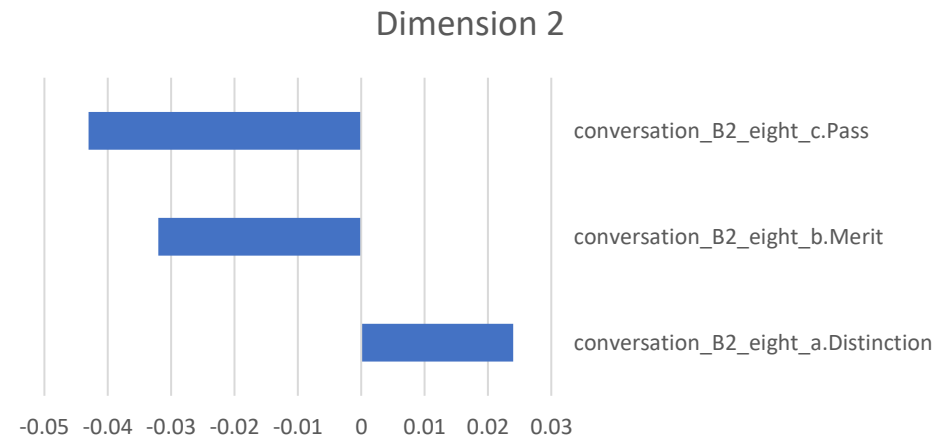
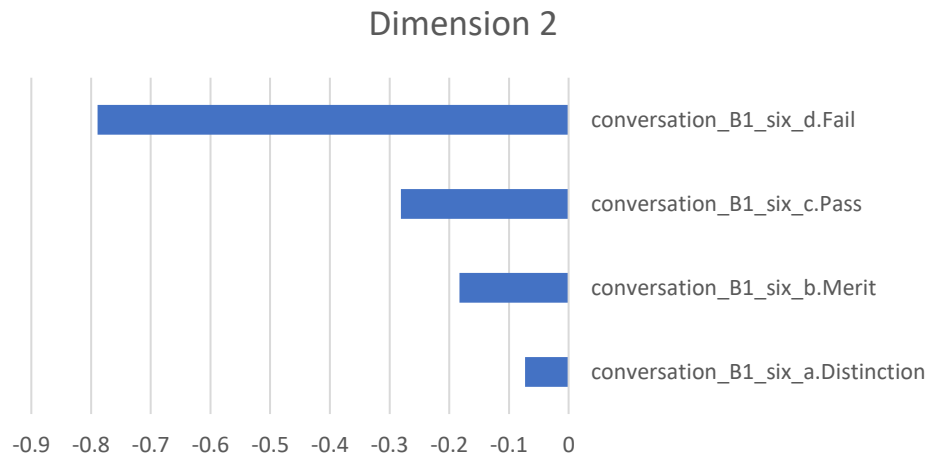
mm it sounds good erm yeah I've never really watched anything similar to that I've haven't watched a lot of cartoons but some things yeah I like to see yeah

Negative: Information seeking

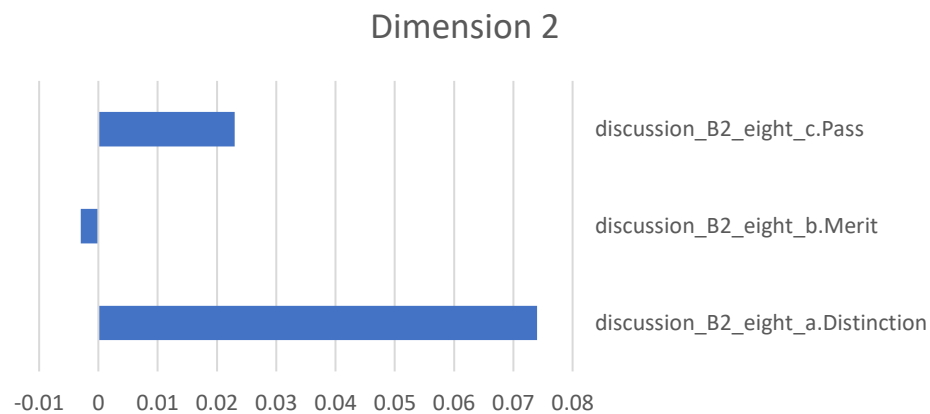
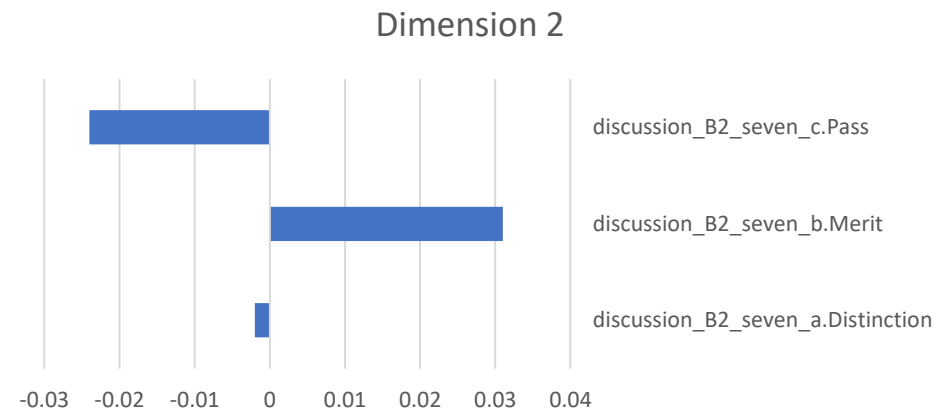
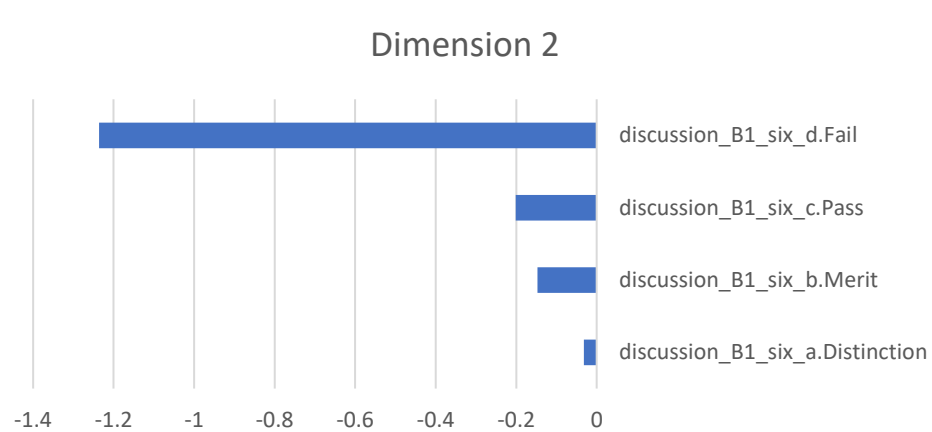
and if you travel to another country what preparations do you have to make?

okay and what did you have to pack before you went? ha= you had a lot of preparation

Learners in the conversation task received a lower overall mark if the examiners produced more turns associated with information seeking as opposed to descriptive turns.



Learners at grade 6 (B1) in the discussion task received a lower overall mark if the examiners produced more turns associated with information seeking as opposed to descriptive and stance turns. No other pattern.



E: yes okay alright then so i-i-if you go to a wedding what what kind of clothes do you wear?

S: if you go to the big s= er st= er shopping centre centre

E: yeah

S: Marineda City

E: yeah

S: are a lot of shops

E: <pause/> <unclear text="just"/> to to buy elegant clothes

S: yes

Learner turns

- Dimension 2: Involved versus Informational
 - Dimension 3: Irrealis (unknown) versus Realis (known)
 - Dimension 4: Infer versus Reveal
 - Dimension 5: Narrative versus Non-narrative
 - Dimension 6: Opinionated Narrative vs. Situation Dependent Commentary
-
- Finding: a narrower repertoire of functions.
 - BUT – a different role in the interaction. Is there any evidence that proficiency may relate to the widening of the repertoire?
 - Look at task, grade and then L1 speakers performing the same tasks.

Dimension 2

Positive: Involved

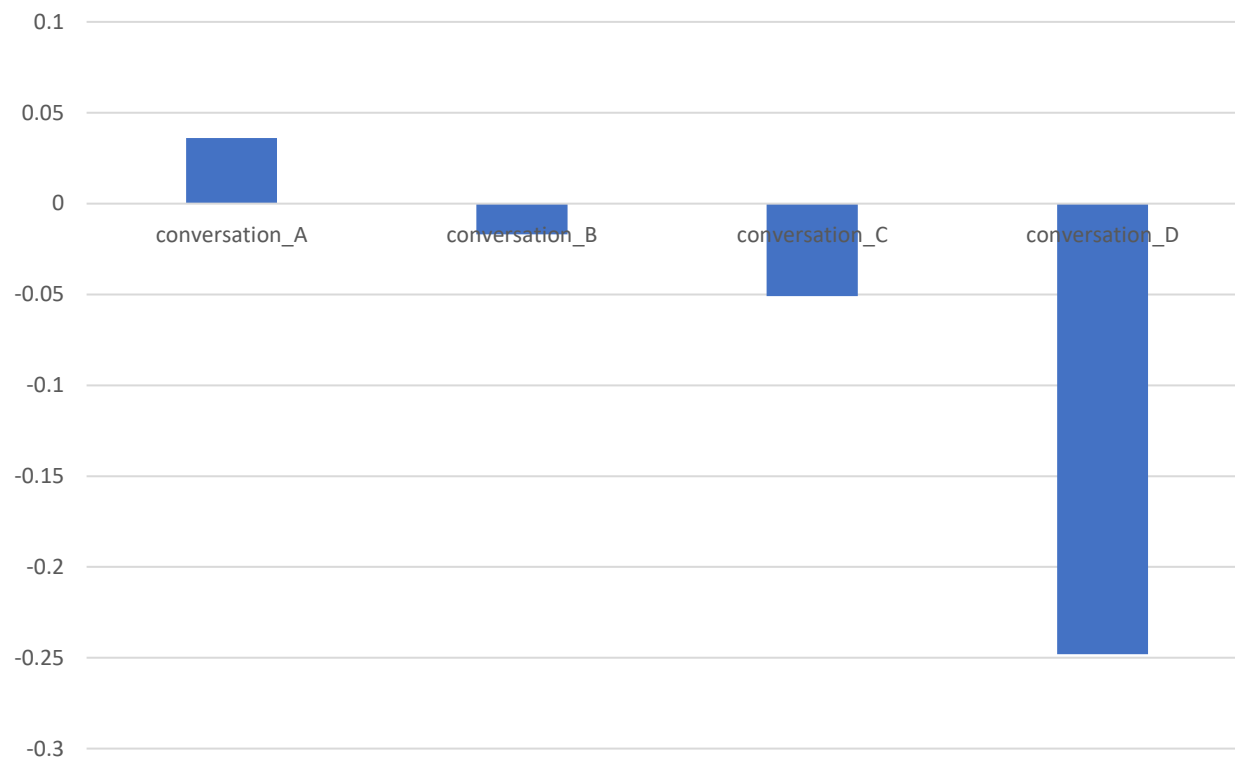
- Copular Verb (she *looks* beautiful)
- Existential *there* (*there* were very cold but sunny days)
- Comparative (she looks *more* beautiful)
- Predicative Adjective (she looks *beautiful*)
- Passive (*it is exported* to a lot of countries...)
- Cause Subordinator (...*because* it's very famous)
- Pronoun IT (*it*)
- Third person singular verb (*looks*)
- Contrastive Conjunction (*but*)
- Demonstrative Pronoun (*that* looks amazing)

Negative: Informational

- Suasive Verb (well now we'll *move* on to the next part)
- Phrasal Verb (you have to ask me questions to *find out* more information)
- Prediction Modal (*we'll* move on)
- Time Adverb (*now*)
- Stance Verb (you *need* to keep the conversation going)
- Infinitive (you need *to keep* the conversation going)
- Nominalisation (*conversation, information*)
- Object Pronoun (you have to ask *me* questions)
- Indefinite Pronoun (I'll tell you *something* then you have to ask)
- Public Verb (you have to *ask* me questions)
- Question Mark (are you ready?)
- Second Person Pronoun (are *you* ready?)

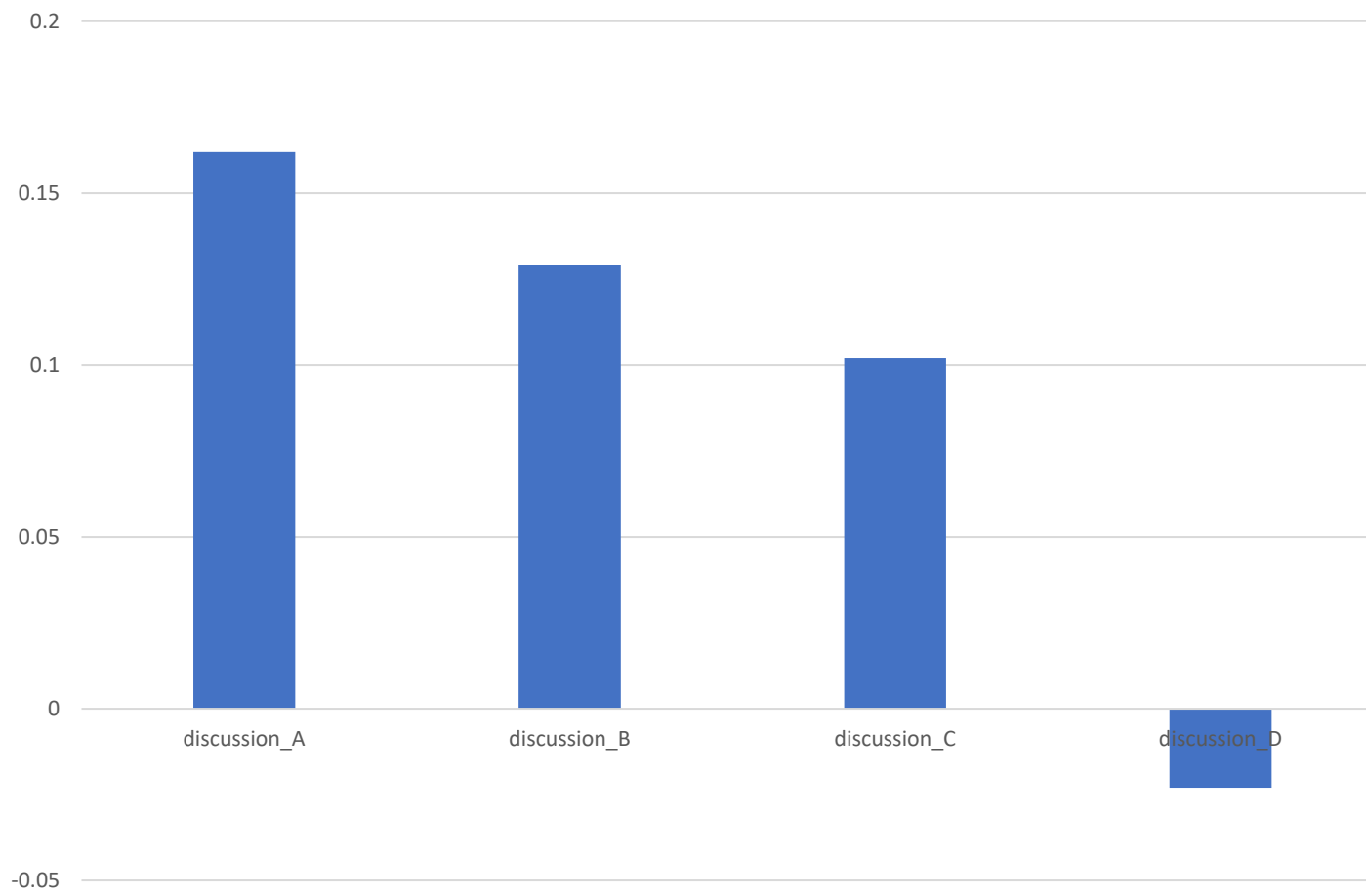
Involved

Dim.2

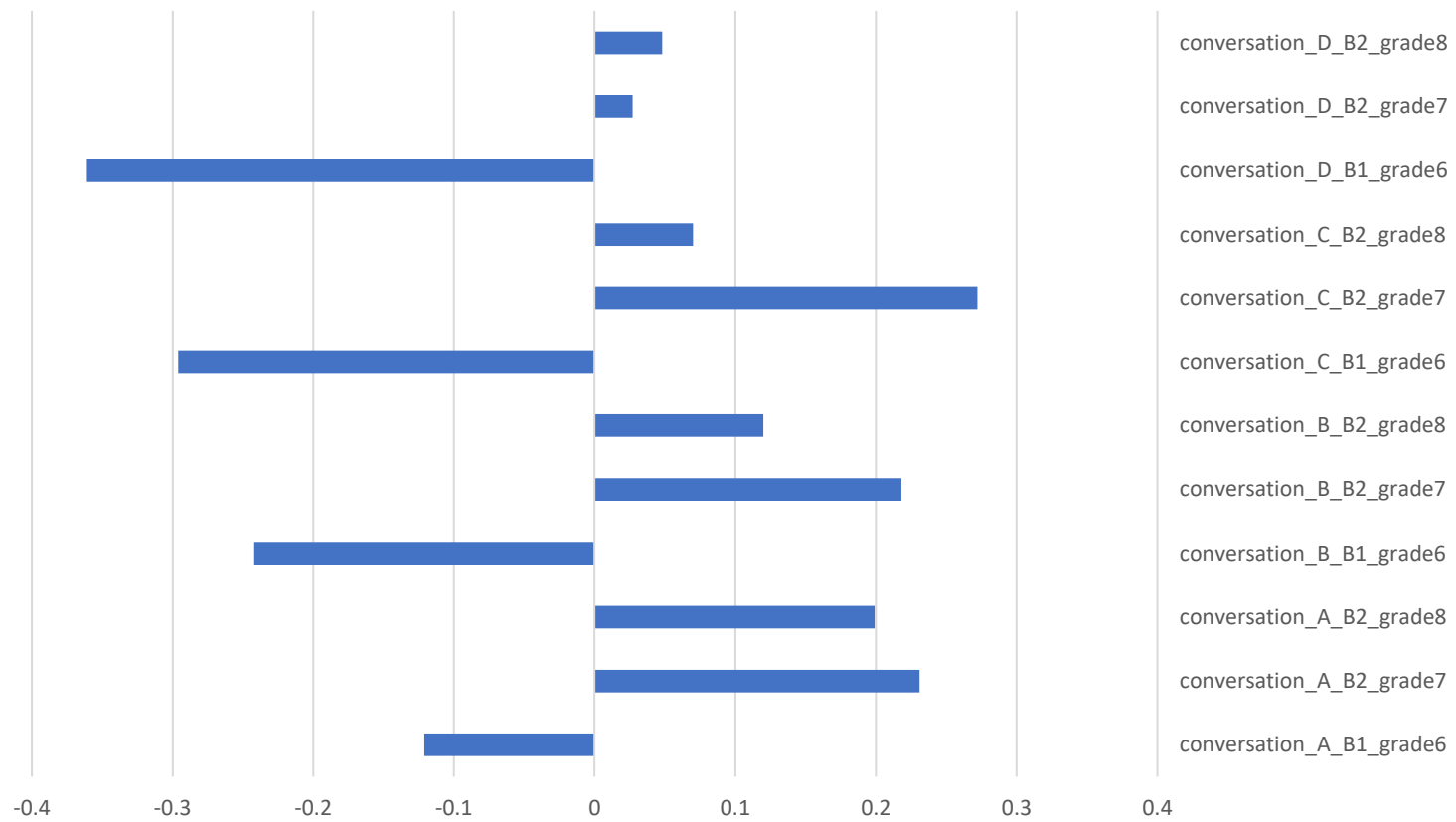


Informational

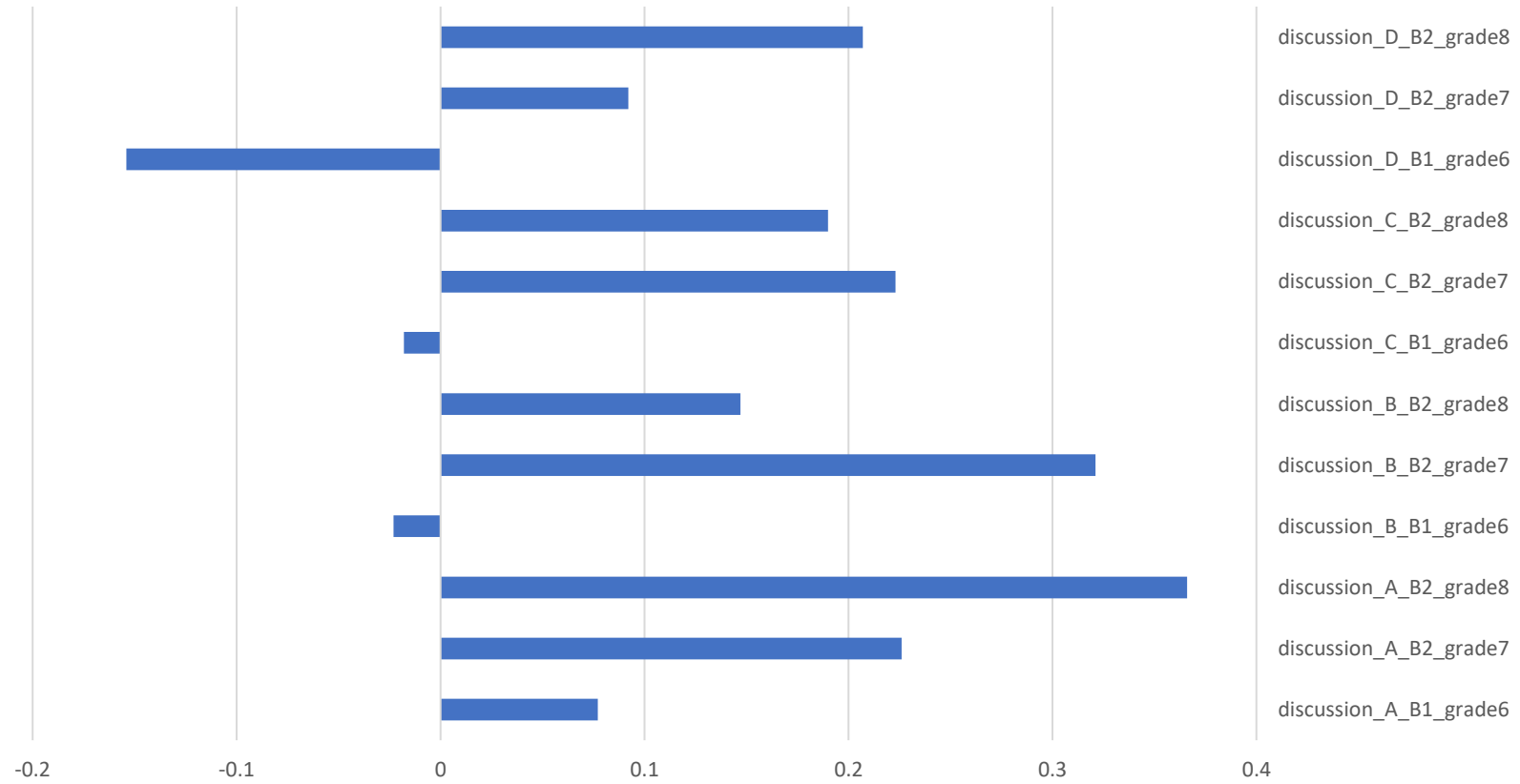
Dim.2



Dim.2



Dim.2



Looking across all three corpora

- To what extent does task determine function at the DU level?
- If the tasks exhaust the repertoire of functions in casual conversation, then the TLC L1 data should show all the functions of DUs, the BNC should be the same
- If the repertoire elicited by the exam is a subset of conversational functions, that shows the focus of the tasks
- If the repertoire elicited by the exam is distinct from conversational English, then this suggests the exam is not a good preparation for everyday interaction.

Functions shared by all three corpora

Function	Dimensions	Shared by	Type
Information seeking	TLC D5-, TLC L1 D5+, BNC D4-	3	All
Informative and instructive	TLC D2-, TLC L1 D3-, BNC D2+	3	All
Seeking and encoding stance	TLC D4-, TLC L1 D4+, BNC D3-	3	All

Functions yet to be acquired by L2 speakers

Function	Dimensions	Shared by	Type
Situation dependent commentary	TLC L1 D5-, BNC D6-	2	L1 Only

Functions elicited in the exam only

Function	Dimensions	Shared by	Type
Informational narrative	TLC D4+, TLC L1 D4-	2	L2 and L1 exam
Irrealis	TLC D3+, TLC L1 D2-	2	L2 and L1 exam
Realis	TLC D3-, TLC L1 D2+	2	L2 and L1 exam

Function	Dimensions	Shared by	Type
Attitudinal descriptions	BNC D2-	1	L1 Conversation
Elaborated speech	BNC D1-	1	L1 Conversation
Informational recounts	BNC D3+	1	L1 Conversation
Narrative	BNC D5-	1	L1 Conversation
Non-narrative	BNC D5+	1	L1 Conversation
Opinionated narrative	BNC D6+	1	L1 Conversation
Reveal	BNC D4+	1	L1 Conversation

L1 Functions deployed in exam, but not seen in everyday conversation

Function	Dimensions	Shared by	Type
Affective	TLC L1 D3+	1	L1 Exam
Extended narrative	TLC L1 D1-	1	L1 Exam

Functions seen in L2 speech only

Function	Dimensions	Shared by	Type
Descriptive and affective	TLC D2+	1	L2 Exam
Persuasion	TLC D5+	1	L2 Exam

Conclusion

- MCA is a powerful technique for exploring sparse data
- It can bridge the micro-macro divide
- Supplementary variables are a great way of exploring the interaction of metadata elements with patterns observed
- Discourse is co-constructed and interlocutors have different roles – we can see this
- Discourse is developing in spoken learner English
- The exam elicits a core of functions typical of everyday language use
- Yet it also elicits functions specific to the context and the tasks in the exam
- There is much more to show and say so Look out for the book!