

The *WriMA corpus* and its contributions to language teaching and material development.

Maria N. Melissourgou

University of the Aegean

2nd International EAP & ESP Conference
University of Crete, Heraklion, 20-23 September 2019




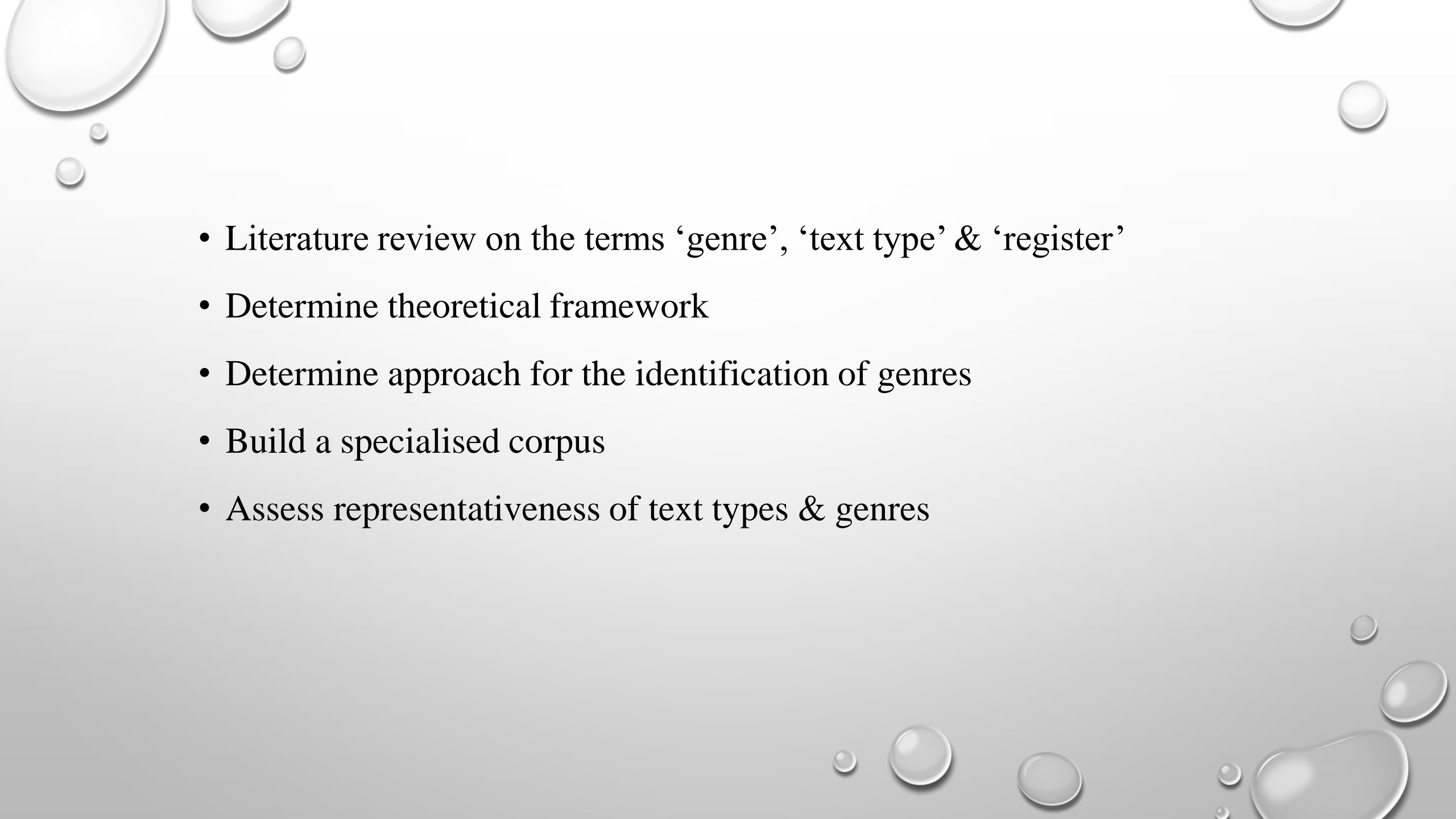


Research question 1

A. What is the representation of text types in model answers offered in English language teaching material?

B. What would this representation be if texts were categorized according to genre instead of text type?



- 
- Literature review on the terms ‘genre’, ‘text type’ & ‘register’
 - Determine theoretical framework
 - Determine approach for the identification of genres
 - Build a specialised corpus
 - Assess representativeness of text types & genres

Methods

- Corpus linguistics & Genre analysis
- The **WriMA** corpus [a pedagogical genre-based corpus of writing model answers]
 - 1151 model writing answers (253,025 tokens)
- 93 different sources
- Sub-corpora initially according to text type
- POS tagged & manually annotated
- Metadata: text source, text type (according to source), rubric, exam, CFR level
- Genre identification method

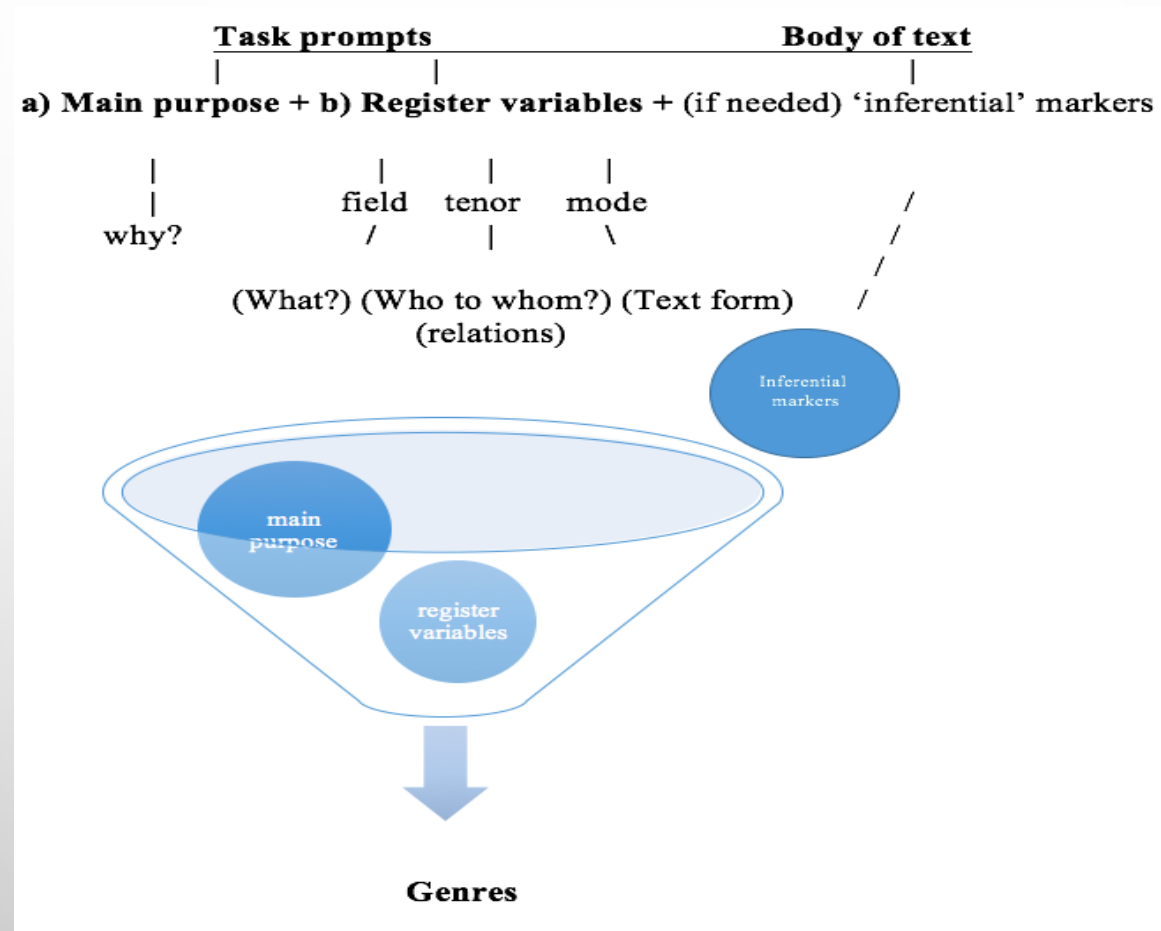


Figure 1. Method for identifying genres from task prompts

Initial corpus categories	# of texts	Initial representation in corpus	Genre categories	# of texts	Genre representation in corpus
Essay	415	36%	Discursive Essay	176	15.3%
			Expository Essay	85	7.4%
			Factorial	31	2.7%
			Consequential	27	2.3%
			Factorial and Consequential Essay	15	1.3%
			Descriptive Essay	58	5%
			Solutions to a problem Essay	23	2%
Report	176	15.3%	Data Report	100	8.7%
			Personal Observation Report	76	6.6%
Article	88	7.6%	Descriptive Article	40	3.5%
			Expository Article	31	2.7%
			Informational Article	9	0.8%
			Solutions to a problem Article	8	0.7%
Review	39	3.4%	Book/Film Review	39	3.4%
Story	78	6.8%	Short Story	78	6.8%
Formal Letter Formal Email	171	14.9%	Complaint Letter/email	47	4.1%
	24	2.1%	Formal Informational Letter/email	38	3.3%
			Application Letter/email	30	2.6%
			Opinion as a response letter/email	28	2.4%
			Reference Letter	12	1%
			Formal invitation Letter/email	8	0.7%
			Making suggestions formal Letter/email	8	0.7%
			Formal Apology Letter/email	6	0.5%
			Formal Letter of Request	6	0.5%
			Solutions to a community problem Letter	5	0.4%
			Nomination Letter	4	0.3%
			Resignation Letter	3	0.3%
Informal Letter Informal Email	105	9.1%	Advice Letter/email	61	5.3%
	55	4.8%	Personal news Letter /email	33	2.9%
			Informal Informational Letter/email	29	2.5%
			Informal invitation Letter/email	27	2.3%
			Informal apology letter	7	0.6%
			Informal letter of request	3	0.3%
Total	1151	100%	Total	1151	100%

Table 1. The representation of text types and genres in English language educational material

ORIGINAL PAPER

Genre Identification Based on SFL Principles: The Representation of Text Types and Genres in English Language Teaching Material

Maria N. Melissourgou¹ · Katerina T. Frantzi¹

Received: 25 November 2016 / Accepted: 14 April 2017
© The Author(s) 2017. This article is an open access publication

Abstract Genre awareness has recently gained a lot of attention due to the recognition of the vital impact of genres in discourse comprehension and production. At the same time, the rapid development of corpus linguistics studies has caused a reconsideration of methodological issues such as the classification of texts during corpus building. This study aims to untangle commonly confusing terms such as ‘text type’, ‘genre’ and ‘register’, reviewing their use by prominent researchers. It then attempts to investigate the range of genres involved in writing tasks presented in English language teaching material. Reporting on experience from the Writing Model Answers corpus classification of texts, we explain how we



Research question 2:

What are the most prevalent features of individual genres?



Methodology

focus on 8 genres
detailed analysis of individual genres

Sub-corpus	# of tokens	# of texts
Expository Essay	24.347	85
Discursive Essay	47.153	176
Descriptive Essay	15.448	58
Personal Observation Report	17.702	76
Data Report	16.828	100
Short Story	15.558	78
Complaint Letter	8.770	47
Advice Letter	9.938	61

Table 2. Genre-based sub-corpora chosen for analysis

- Genre analysis → Rhetorical moves/stages
Lexicogrammatical features
- Corpus analysis → Keywords
Keyword lists & keyness scores
Collocations
Metadata

Papers from the 10th Lancaster University
Postgraduate Conference in Linguistics &
Language Teaching 2015

Volume 10

The language of Reports in general English language testing:
A corpus-based analysis.

Maria N. Melissourgou

University of the Aegean (Greece)

Abstract

Candidates of international English language exams are often asked to write a report. However, this kind of report is different from the frequently appearing research reports or reports in professional contexts. Information to guide learners specifically about this genre is scarce and not well documented. Therefore, the aim here, is to provide empirical data on the qualities competent writers use to achieve the communicative purpose of this genre. Having previously identified two basic variations of reports in this context, the Data Report and the Personal Observation Report, I describe the most prominent features of the language used in personal view reports based on quantitative evidence. Using a sub-corpus of the WriMA (Writing Model Answers) corpus (Melissourgou & Frantzi, 2015) I conduct a genre and corpus analysis. The WriMA corpus is a genre-based corpus especially created for this type of research. Examining frequent common words, keywords and



LIF – Language in Focus Journal, Volume 2, 2016, DOI: 10.1515/lifjsal-2016-0009

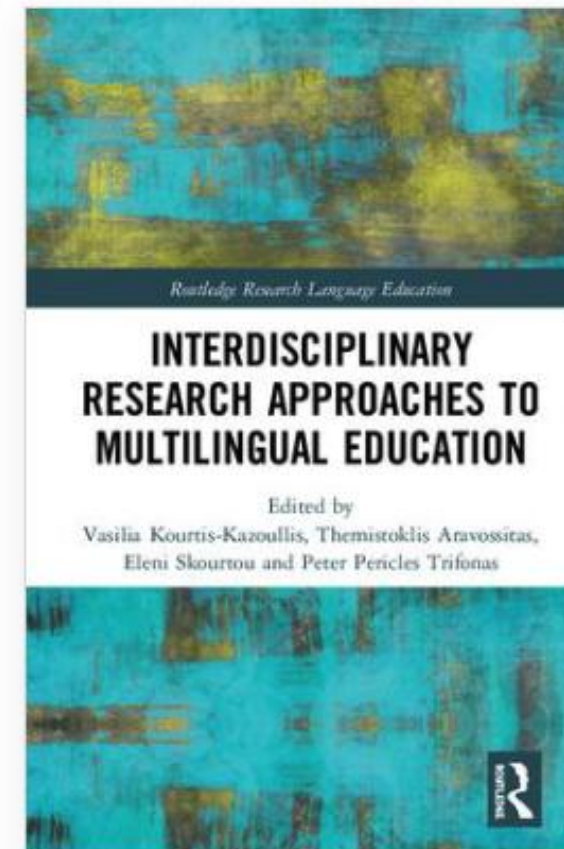
Register analysis based on corpus evidence: the case of Reports in IELTS

Maria N. Melissourgou, Katerina T. Frantzi

University of the Aegean, Greece

Abstract

The IELTS exams attract numerous candidates globally and are considered high-stakes exams due to the impact the results can have on the candidates. The need therefore for goal-oriented guidance is crucial. Unfortunately, this preparation is based on teachers' intuition on what is common and appropriate. This paper describes the typical lexicogrammatical features of Reports based on objective quantitative evidence. The WriMA (Writing Model Answers) corpus consists of model writings from preparation material. The DataReports sub-corpus used




Chapter 10
Stories, Second Language Writers
and Genre Awareness



Research question 3:

a. How are genres in this context related to each other?

B. Is the text type classification of model texts justified? Is the implied similarity within such categories stronger than with texts in other categories?



Method

- ❖ set of 16 features
- ❖ features linked to text properties
- ❖ measuring each feature for each genre
- ❖ distance between pairs of genres based on each feature

Features measured		Association with text properties	
1.	Passive voice verbs	1.	Objectivity and formality
2.	Lexical density	2.	Informational density
3.	Nominalisation	3.	Objectivity and elaboration
4.	STTR	4.	Lexical variety
5.	Word length	5.	Lexical complexity
6.	Words per sentence	6.	Syntactic complexity
7.	3-word lexical bundles	7.	Standardised language
8.	Verbs (active, present simple)	8.	Reference to timeless present
9.	Verbs (active, past simple)	9.	Reference to past
10.	Connectors—temporal	10.	Events set in time
11.	Connectors—adding	11.	Addition
12.	Connectors—Contrasting	12.	Contrast
13.	Connectors—Causal/consequential	13.	Causality
14.	Modals	14.	Stance
15.	1st person sing. pronoun	15.	Involvement
16.	2nd person sing. pronoun	16.	Reader engagement

Table 3. Features and associated text properties

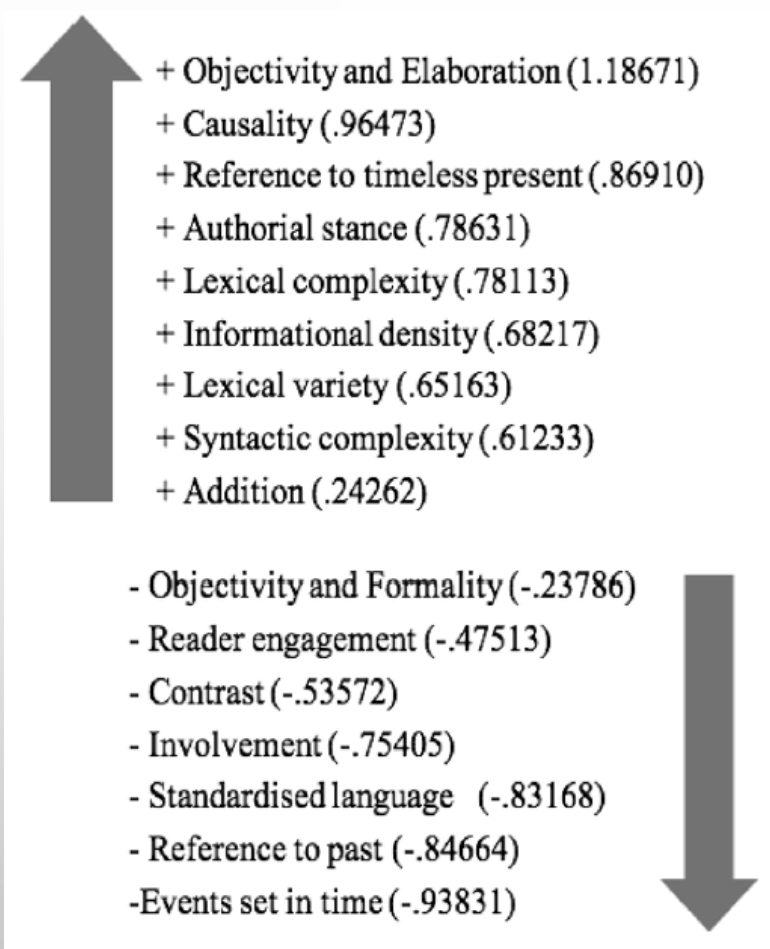


Figure 2. The expository essay

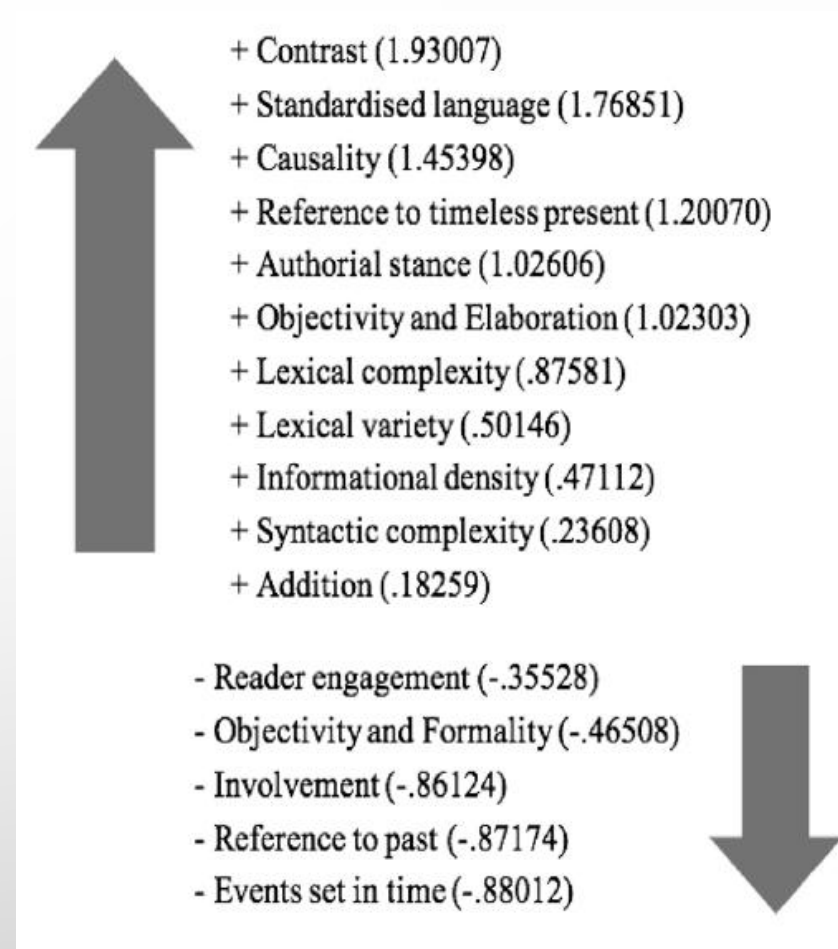


Figure 3. The discursive essay

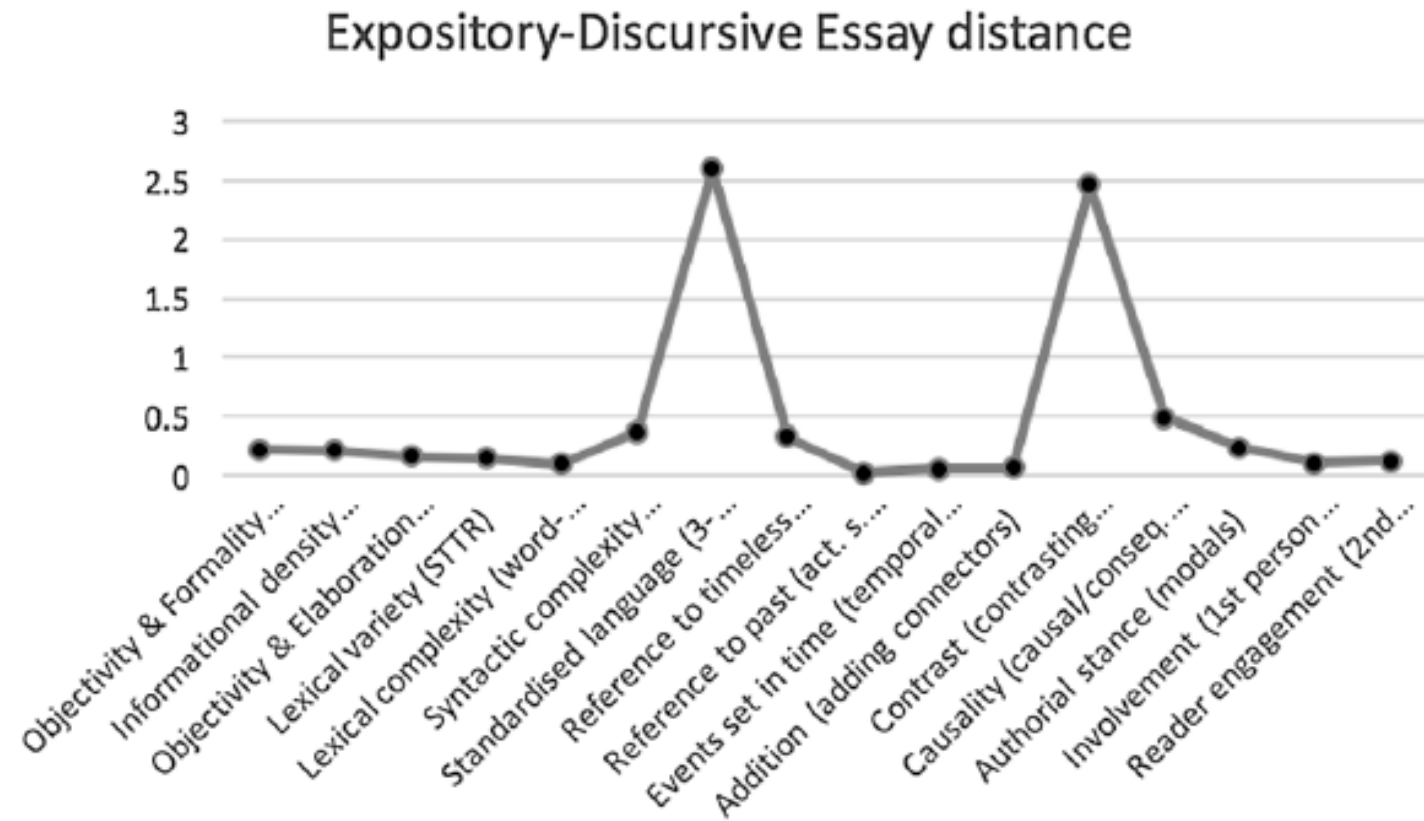


Figure 4. The expository-discursive essay distance

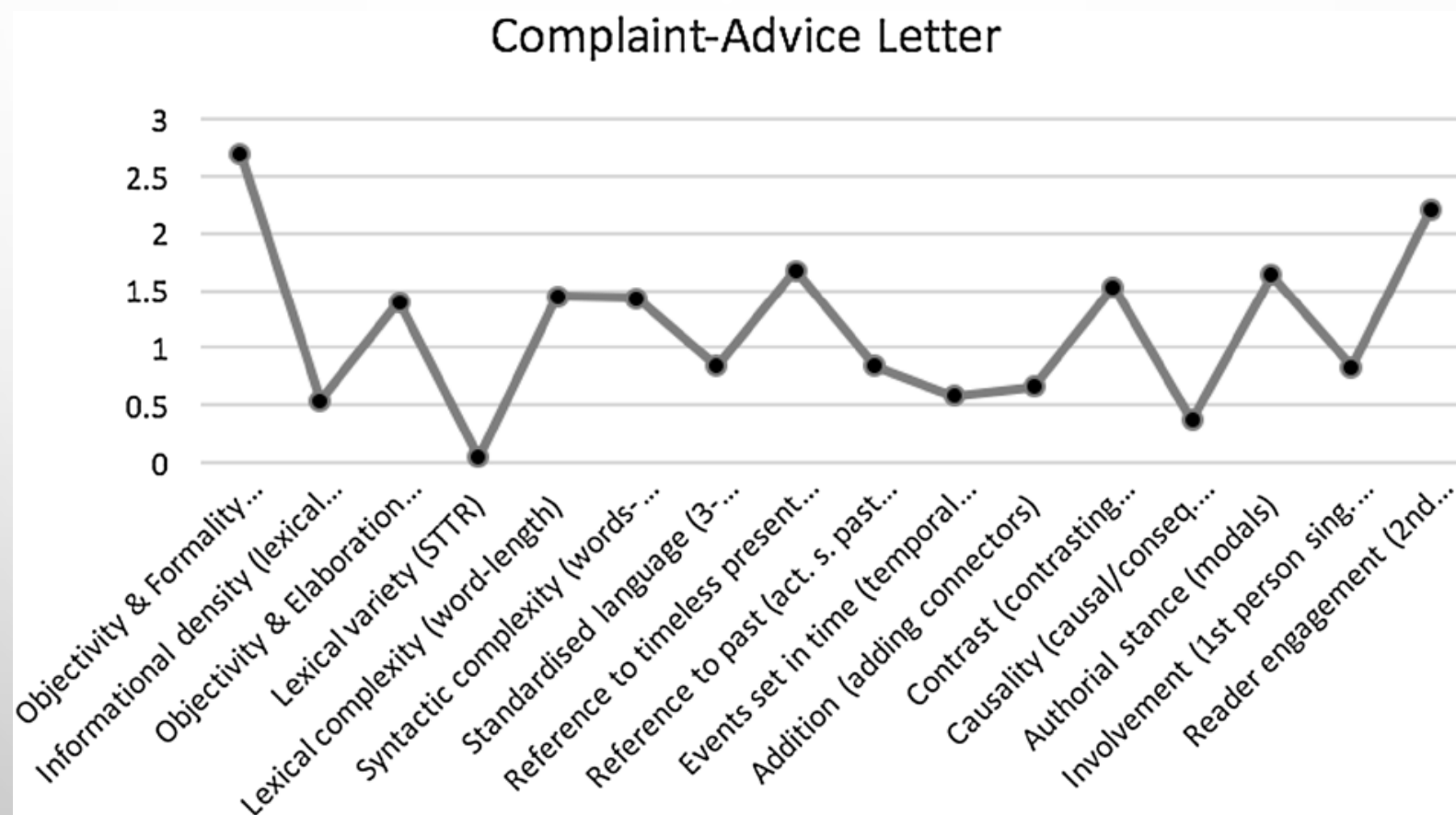


Figure 5. The complaint-advice letter distance

ORIGINAL PAPER

Moving Away from the Implicit Teaching of Genres in the L2 Classroom


Maria N. Melissourgou¹  · Katerina T. Frantzi¹

Received: 30 November 2017 / Accepted: 28 March 2018
© The Author(s) 2018

Abstract English language proficiency exams are often associated with high-stakes decisions. Guidance however, concerning the writing tasks is often implicit. The emphasis is usually placed upon grammatical and lexical features rather than the pragmatic aspects that differentiate the tasks. Aiming to boost genre awareness as part of L2 pragmatic competence in this context the present paper provides a


Conclusions


Review of the terms ‘genre’, ‘text type’ & ‘register’ - lack of consensus –
Positioning on the use of the terms – Genre-identification Method



A great variety of genres have been observed in exam preparation textbooks.


Some genres are overrepresented in textbooks such as Essays but most of the genres are underrepresented.






Typical features of genres can be revealed by using corpus linguistics methods and a genre-based corpus.


Detailed descriptions of genres can make teaching more explicit and less confusing for novice writers.





Relations between pairs of genres can sometimes be surprising. Learners preparing for writing tasks grouped according to text types are not well-equipped to distinguish among genres

Similarity and difference among genres can be explained in the classroom. Indicating the link between lexicogrammatical features and text properties can lead to more conscious writing.



References

Anthony, L. (2014). *TagAnt* (version 1.1.2) [Computer Software]. Tokyo: Waseda University. Available from <http://www.laurenceanthony.net/>

Melissourgou, M. N. & Frantzi, K. T. (2019). Stories, second language writers, and genre awareness. In V. Kourtis-Kazoulis, T. Aravossitas, E. Skourtou and P.P. Trifonas (eds.) *Interdisciplinary Research Approaches to Multilingual Education* pp. 117-129. Abingdon/NY: Routledge.

Melissourgou, M. N. & Frantzi, K. T. (2018). Moving Away from the Implicit Teaching of Genres in the L2 Classroom. *Corpus Pragmatics*, 2(4), 351-373.

Melissourgou, M. N., & Frantzi, K. T. (2017a). Register analysis based on corpus evidence: the case of Reports in IELTS. *Language in Focus*, 2(2), 70–87.

Melissourgou, M. N., & Frantzi, K. T. (2017b). Genre Identification Based on SFL Principles: The Representation of Text Types and Genres in English Language Teaching Material. *Corpus Pragmatics*, 1(4), 373–392.

Melissourgou, M. N. (2016). The language of reports in general English language testing: a corpus-based analysis. In E. Nichele, D. Pili-Moss, & C. Sitthirak (Eds.), *Papers from the 10th Lancaster University Postgraduate Conference in Linguistics & Language Teaching 2015* (pp. 33–52). Lancaster, UK: Lancaster University.

Scott, M., (2015). *Wordsmith tools version 6*, stroud: Lexical Analysis Software. available at <http://lexically.net/wordsmith>

