

University of Leipzig, 1. December 2020

The Enhanced Shakespearean Corpus: Theoretical and practical explorations in Early Modern English

Isolde van Dorst & Sean Murphy

Department of Linguistics and English Language, Lancaster University

Vienna University of Economics and Business

Universitat de Barcelona / Universitat Ramon Llull, Blanquerna

 @ShakespeareLang

<http://wp.lancs.ac.uk/shakespearelang/>



Arts & Humanities
Research Council



THE QUEEN'S
ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2015



Outline

- Introduction to the project
- Two studies:
 - Hardie A, van Dorst I. A survey of grammatical variability in Early Modern English drama. *Language and Literature*. 2020;29(3):275-301. doi:[10.1177/0963947020949440](https://doi.org/10.1177/0963947020949440)
 - Murphy S, Culpeper J, Gillings M, Pace-Sigge M. What do students find difficult when they read Shakespeare? Problems and solutions. *Language and Literature*. 2020;29(3):302-326. doi:[10.1177/0963947020949441](https://doi.org/10.1177/0963947020949441)

The Encyclopaedia of Shakespeare's Language project

(<http://wp.lancs.ac.uk/shakespearelang>)

Main people: Jonathan Culpeper, Dawn Archer, Alison Findlay, Andrew Hardie, Paul Rayson, Jane Demmen, Sean Murphy, Isolde van Dorst, Mathew Gillings, Jakob Ladegaard, Ross Deans Kristensen-McLachlan

Funder: UK's Arts and Humanities Research Council (AHRC)

Timeframe: 2016-ongoing

Aim: Produce the first systematic and comprehensive account of Shakespeare's language using methods derived from corpus linguistics – an approach that uses computers in large-scale language analysis.

A distinctive feature: It's comparative.

What will the project produce?

- The encyclopaedia:
 - First and second volume: dictionary
 - Third volume: plays and characters
 - Fourth volume: social networks
 - Fifth volume: semantic fields and themes
- To be published by Bloomsbury in paper and electronically
- Electronic resources
 - Including the Enhanced Shakespearean Corpus
 - MOOC
 - Searchable database

Enhanced Shakespearean Corpus (ESC)

- Three components:
 - ESC: First Folio Plus (ESC: Folio)
 - ESC: Comparative Plays (ESC: Comp)
 - ESC: EEBO-TCP Segment (ESC: EEBO)
- Additionally:
 - ESC: Quartos
 - ESC: Verse

All accessible through CQPweb:
www.cqpweb.lancs.ac.uk

ESC: Folio

- Core data: The Shakespeare canon
 - All 36 plays of the First Folio (1623)
 - *Pericles: Prince of Tyre* (Quarto 1, 1609)
 - *Two Noble Kinsmen* (Quarto 1, 1634)
- Total of approximately 1 million words
- Key enhancements:
 - Spelling regularisation – to enhance searchability without losing original spellings
 - Grammatical (part-of-speech) tagging – e.g., to facilitate dictionaries and grammars
 - Social annotation – e.g., to facilitate the study of speech styles
- Rich meta-data:
 - Speaker: ID, gender, social status
 - Text: play ID, genre, scene, accent/dialect/language

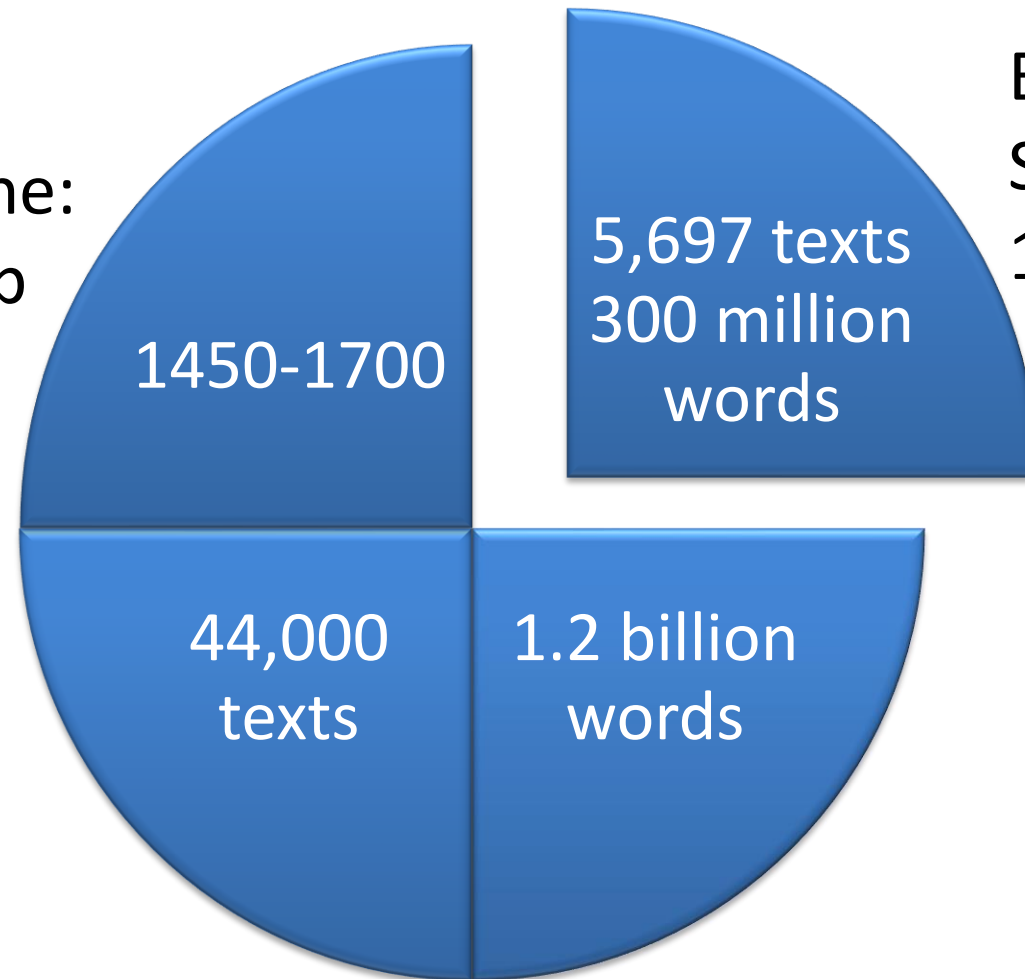
ESC: Comp

-
- Built to contextualise Shakespeare's language (relative to that of a group of his peers)
 - Designed to mimic the size and genre composition of the ESC: Folio
 - Approximately 1 million words
 - 46 plays by 24 other playwrights from the same time period as Shakespeare
 - Same enhanced tagging done on this corpus as on ESC: Folio

ESC: EEBO

EEBO-TCP

Early English Books Online:
Text Creation Partnership



EEBO-TCP
Segment
1560-1640

A survey of grammatical variability in Early Modern English drama

- Andrew Hardie & Isolde van Dorst
- Investigation of grammatical features to understand their variability in Early Modern English drama
- And the extent to which Shakespeare's grammatical style is distinct from or similar to that of his contemporaries
- Using the ESC: Folio and the ESC: Comp

Method

- 15 grammatical features of stylistic interest
 - Leech and Short (1981) → stylistic analysis of grammatical features
 - Biber (1988) → multi-dimensional analysis
 - Overlapping features: *that* clauses, *wh*-clauses, infinitives, past tense verbs, perfect aspect verbs
- Between the narrow focus of close reading and the naïve quantitative metrics of authorship analysis
- Using both statistical comparison and visual representation of variability

15 features

Simple features (6)	Complex features (9)	
Verbs	Paste tense verbs	
Nouns	Perfect aspect verbs	Verb <i>have</i> , followed by past participle
Adjectives	Present tense verbs	
Adverbs	1st person pronouns	
Pronouns	2nd person pronouns	
Determiners	3rd person pronouns	
	<i>That</i> -clause verb complements	Complementiser <i>that</i> (conjunction), plus surrounding possible patterns <i>"They said that Palamon had Arcite's body"</i> (TNK)
	Infinitives	Word tagged as TO (infinitive marker) <i>"To be or not to be"</i> (Ham)
	<i>Wh</i> -clause verb complements	Wh-word followed by a clause and preceded by a verb <i>"Porter, remember what I gave in charge"</i> (1H6)

Statistical interpretation

Feature	Effect	Interpretation
Verbs	0.98	
Nouns	0.94	Shakespeare's style disprefers nouns .
Adjectives	1.01	
Adverbs	0.97	
Pronouns	1.05	Shakespeare's style prefers pronouns .
Determiners	0.95	Shakespeare's style disprefers determiners .
Past tense verbs	1.06	Shakespeare's style prefers past tense .
Perfect aspect verbs	1.15	Shakespeare's style prefers perfect aspect .
Present tense verbs	0.97	
1st person pronouns	1.00	Frequency of 1st person pronouns appears not to discriminate Shakespeare's style from other playwrights.
2nd person pronouns	1.00	Frequency of 2nd person pronouns appears not to discriminate Shakespeare's style from other playwrights.
3rd person pronouns	1.05	Shakespeare's style prefers 3rd person pronouns .
<i>That</i> -clause verb complements	1.33	Shakespeare's style strongly prefers this type of clause as verb object .
Infinitives	0.87	Shakespeare's style disprefers subordination via to-infinitive clauses .
<i>WH</i> -clause verb complements	0.87	Shakespeare's style disprefers this type of clause as verb object .

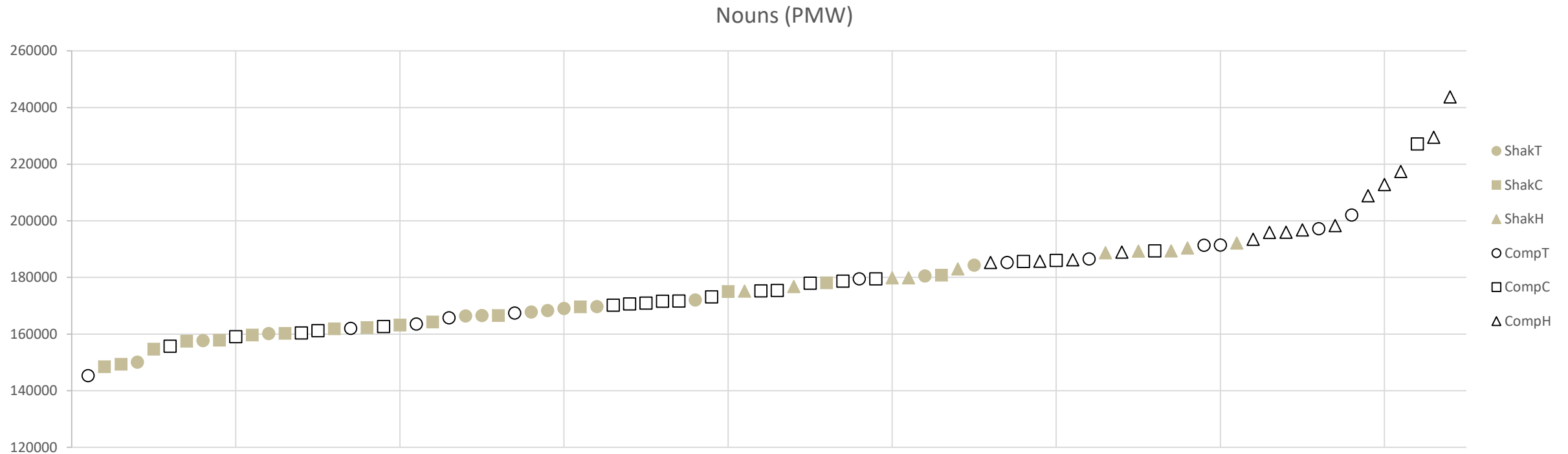
- Effect sizes (RRF to 2 d.p.) for differences in grammatical feature use between Shakespeare and other playwrights.
- Less nouns/less determiners + more pronouns → less informationally dense and more speech-like
- Less infinitives + more *that*-clauses → could again suggest lower information density
- More past/perfect aspect verbs + 3rd pronoun → higher narrativity
- In sum: Shakespeare's grammatical style is differentiated by its narrativity and preference for lesser information density

Variability

- Statistical measures have definitely shown some differences in grammatical style preferences, but do not account for variability in any way (e.g., genre, proportions of prose/poetry)
- Rather than applying more advanced statistical measures, we decided to examine the actual underlying data visually to understand what the variation is doing.

Nouns

- Statistics: Shakespeare's grammatical style disprefers nouns



- Actually: Tragedies and comedies disprefer nouns → just a little stronger in Shakespeare's style
- Histories are usually information dense

Conclusions

- Shakespeare's grammatical style prefers features of narrativity → third person pronouns, past tense verbs, perfect aspect verbs
- Shakespeare's grammatical style disprefers features commonly associated with information density → nouns and determiners
 - Often seen as a feature of orality
 - ...and is, independent of playwright, associated with comedies/tragedies, not histories
- First and second pronouns (also linked to orality) are no more preferred in Shakespeare's grammatical style than in the average style of his contemporaries
- Main significance: this study demonstrates that EME grammar and grammatical style can be analysed quantitatively using corpus data



Encyclopedia of
Shakespeare's
Language Project



University of Leipzig, 1st Dec 2020

The Enhanced Shakespearean Corpus: Theoretical and practical explorations in Early Modern English (2)

Difficulties reading Shakespeare? Problems and solutions

Sean Murphy

@ShakespeareLang



Arts & Humanities
Research Council



THE QUEEN'S
ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2015



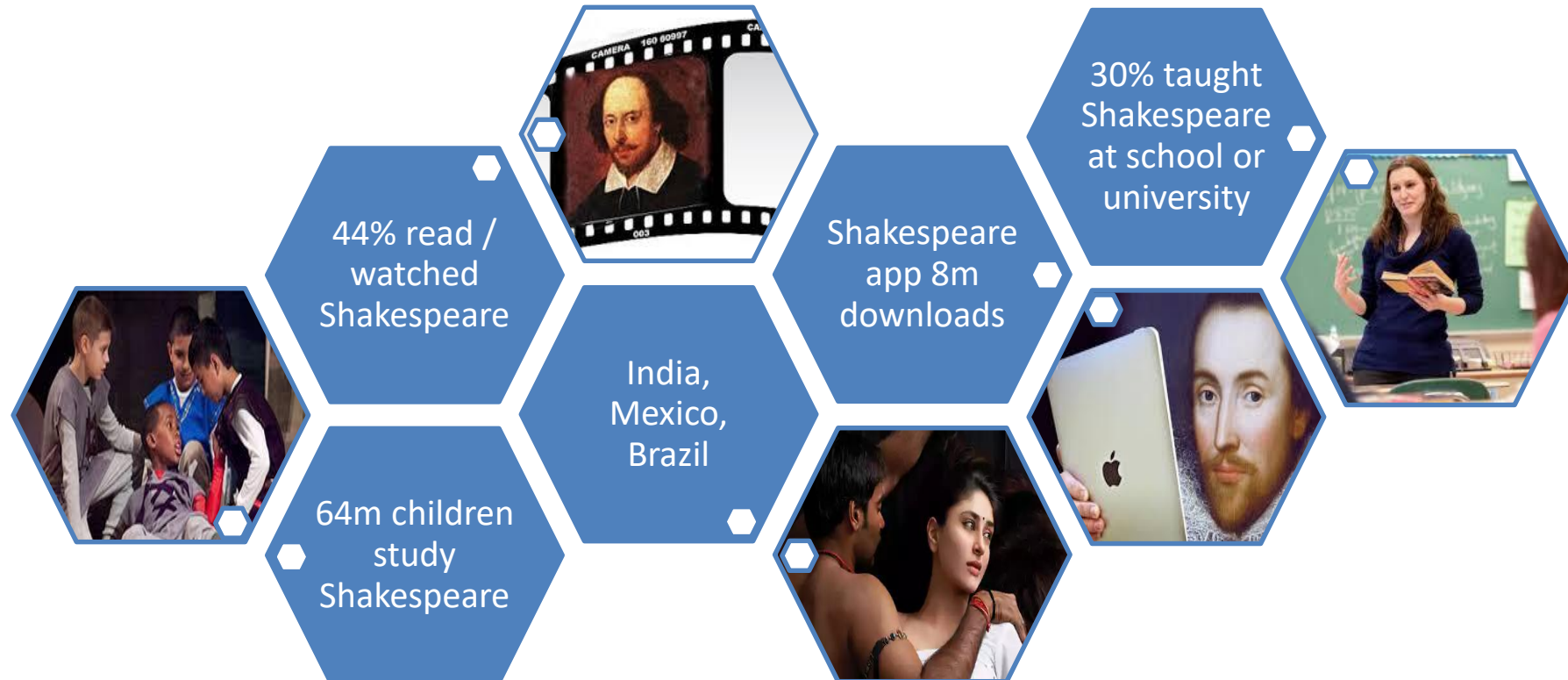
Outline

Shakespeare - teaching
- some issues

a research study: what
do students find
difficult when reading
Shakespeare?

an inductive corpus
approach to
discovering
Shakespeare's language

Shakespeare's influence abroad: 18,000 people / 15 countries. (Donaldson, 2016)



Approaches to teaching Shakespeare

Textual

- close reading
- simplified texts (*No Fear Shakespeare*)

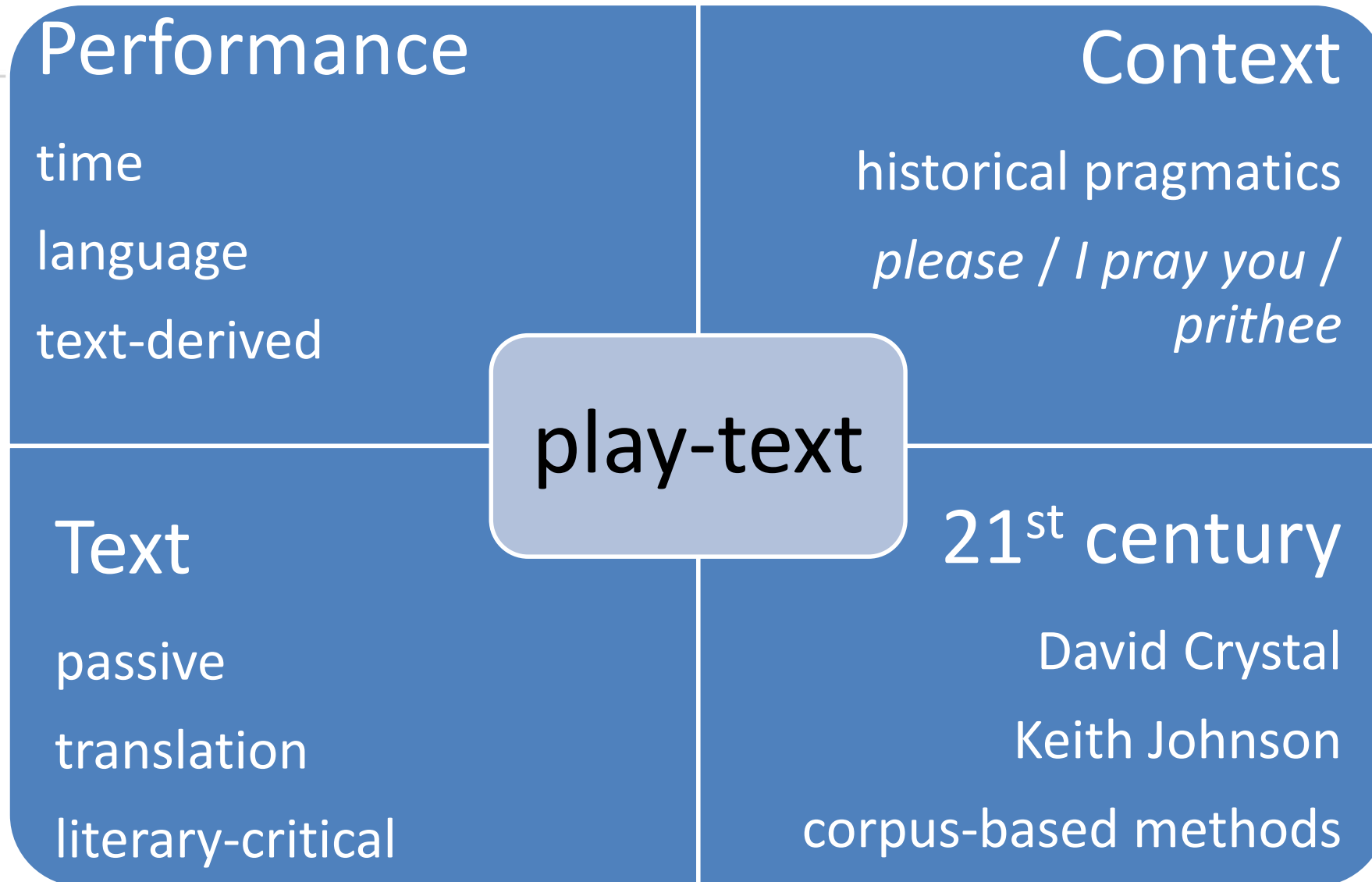
Contextual

- historicism and cultural materialism
- Sir Thomas More addresses anti-immigration mob

Active

- plays understood and experienced through dramatic realisation

Some issues



Participants

Lancaster - Barcelona - Joensuu

English as a first language



Linguistics and literature
undergraduates

English as an additional language



UNIVERSITY OF
EASTERN FINLAND

English Studies undergraduates
/ Language and Culture
undergraduates

Task 1: Extracts

Example

Who would fardels bear, *something heavy?*
To grunt and sweat under a weary life

Play	Character(s)	Extract (V: verse; P: prose)
<i>Richard III</i> (1.2.230-240)	Richard	V: Was ever woman in this humour wooed? [...] And yet to win her! All the world to nothing!
<i>Taming of the Shrew</i> (4.1.177-185)	Petruchio	V: Thus have I politicly begun my reign [...] That bate and beat and will not be obedient.
<i>Henry V</i> (3.6.117-124)	Mountjoy	P: Thus says my king: Say thou to Harry of England [...] see his weakness, and admire our sufferance.
<i>Hamlet</i> (5.1.15-20)	First Clown	P: Give me leave. Here lies the water [...] his own death shortens not his own life.
<i>King Lear</i> (1.1.94-105)	Lear/Cordelia	V: How, how, Cordelia! Mend your speech a little [...] But goes thy heart with this?
<i>Much Ado About Nothing</i> (3.3.19-26)	Dogberry	P: Well, for your favour, sir, why, give God thanks [...] you are to bid any man stand, in the Prince's name.

Task 2: words/phrases

① I don't understand it at all ② I don't understand it very well ③ I'm not sure if I understand it ④ I understand it quite well ⑤ I understand it very well

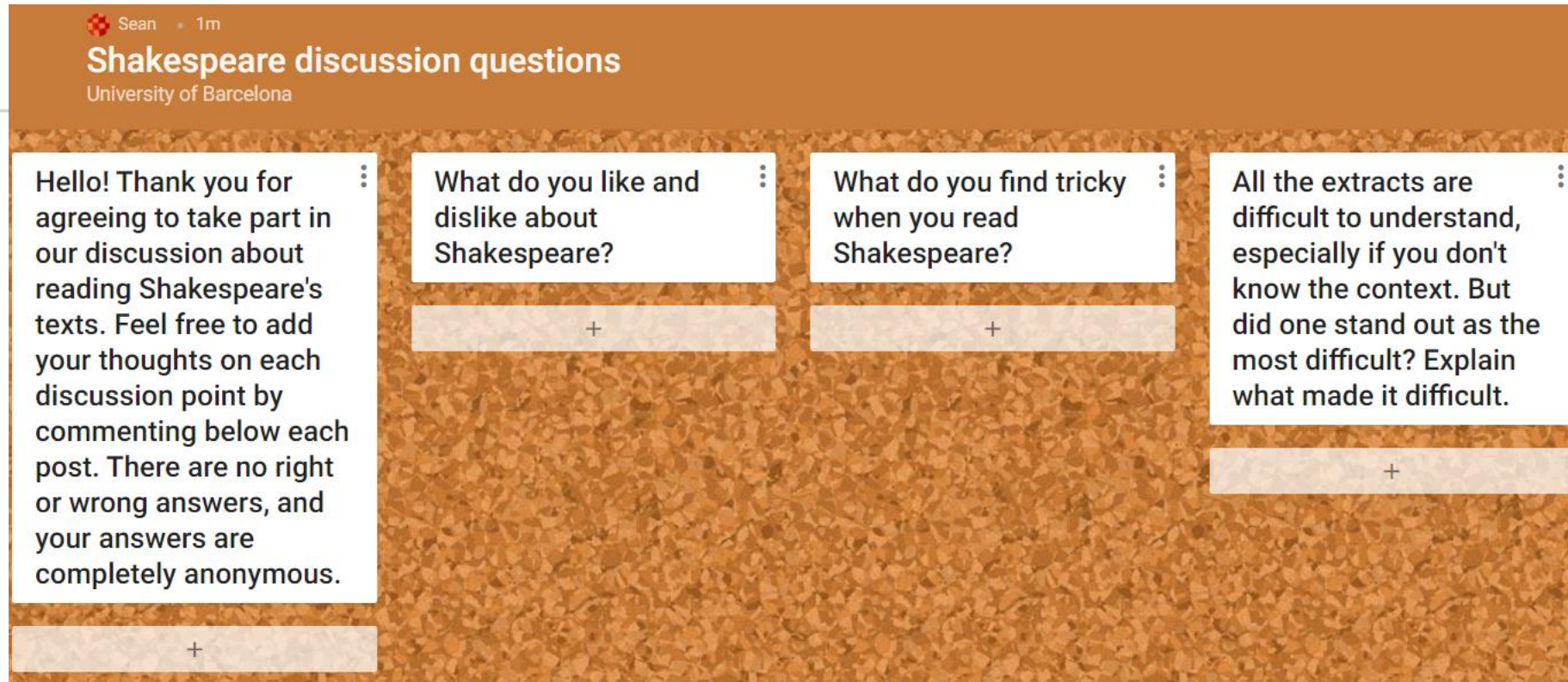
Play	Item	Aspect of language
H5	The vasty fields of France	metrical convenience
Ham	Woo't weep? Woo't fight? Woo't fast?	archaic vb. + contraction
TN	no woman has, nor never none	multiple negation
LLL	Judas I am, yclept Maccabaeus.	archaic verb
TNK	You must e'en take it patiently.	contraction
MV	Let his deservings, and my love withal	archaic adv./prep.
Cym	I have assailed her with musics	plural (now sing.)
1H4	I tell these news to thee?	concord
KL	a better where to find	functional shift
1H6	'Twas time, I trow , to wake	archaic verb

Task 2: Sentences (cont.)

① I don't understand it at all ② I don't understand it very well ③ I'm not sure if I understand it ④ I understand it quite well ⑤ I understand it very well

Play	Item	Aspect of language
Wiv	Bless thee, bully doctor!	false friend
TA	Gramercy , lovely Lucius	borrowing from French
2H4	whose chin is not yet fledged	coinage
Cor	My words disbenched you not	affixation
AYL	I would fain see this meeting	archaic adverb
JC	the people fell a-shouting	a- as particle
Tem	Well demanded, wench	archaic noun
AYL	a properer man	comparative
Tem	The mistress which I serve	grammatical change
RJ	So early walking did I see your son	word order

Discussion on Padlet



Sean · 1m

Shakespeare discussion questions

University of Barcelona

Hello! Thank you for agreeing to take part in our discussion about reading Shakespeare's texts. Feel free to add your thoughts on each discussion point by commenting below each post. There are no right or wrong answers, and your answers are completely anonymous.

What do you like and dislike about Shakespeare?

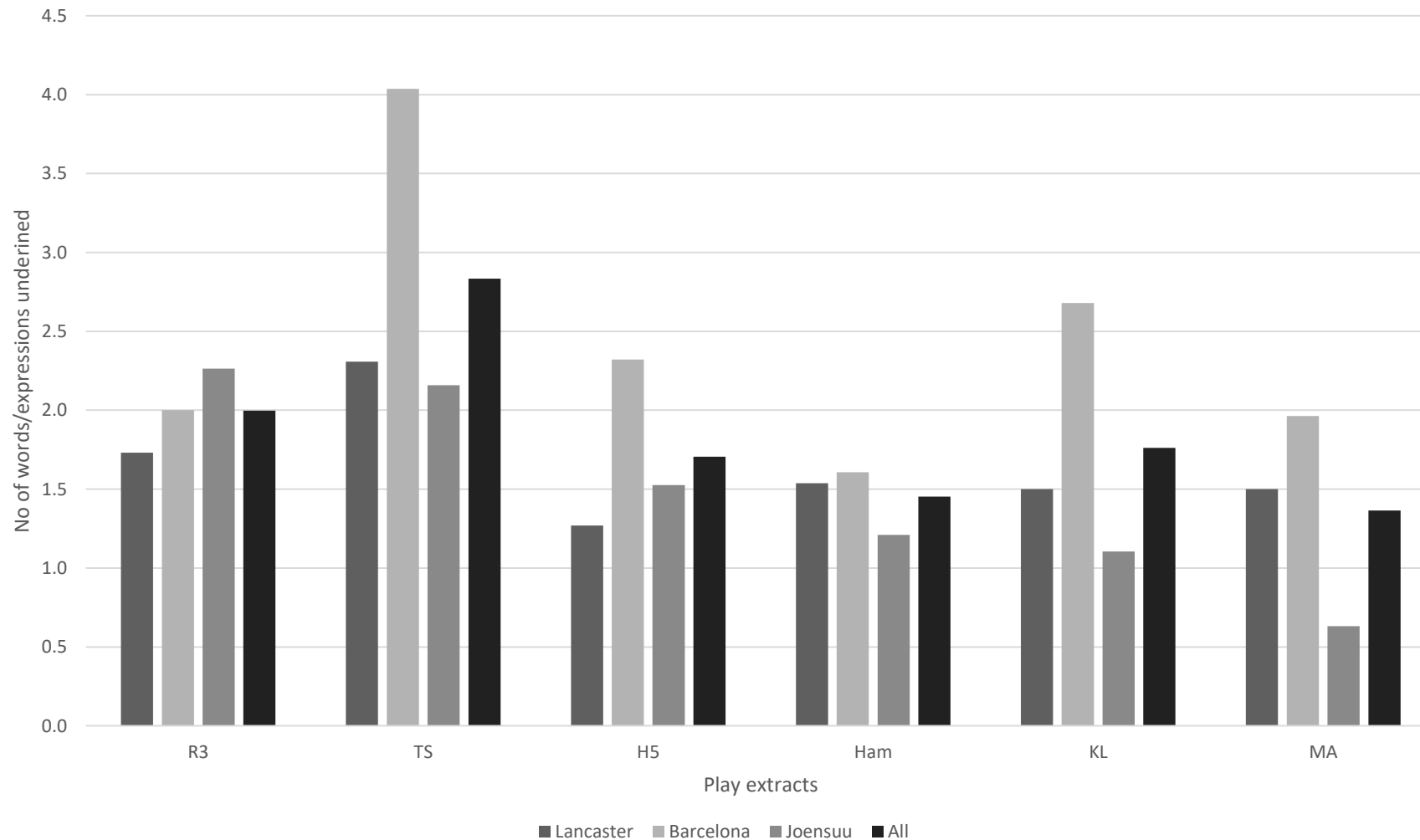
What do you find tricky when you read Shakespeare?

All the extracts are difficult to understand, especially if you don't know the context. But did one stand out as the most difficult? Explain what made it difficult.

To access this padlet from your tablet or phone, scan this code from our mobile app.



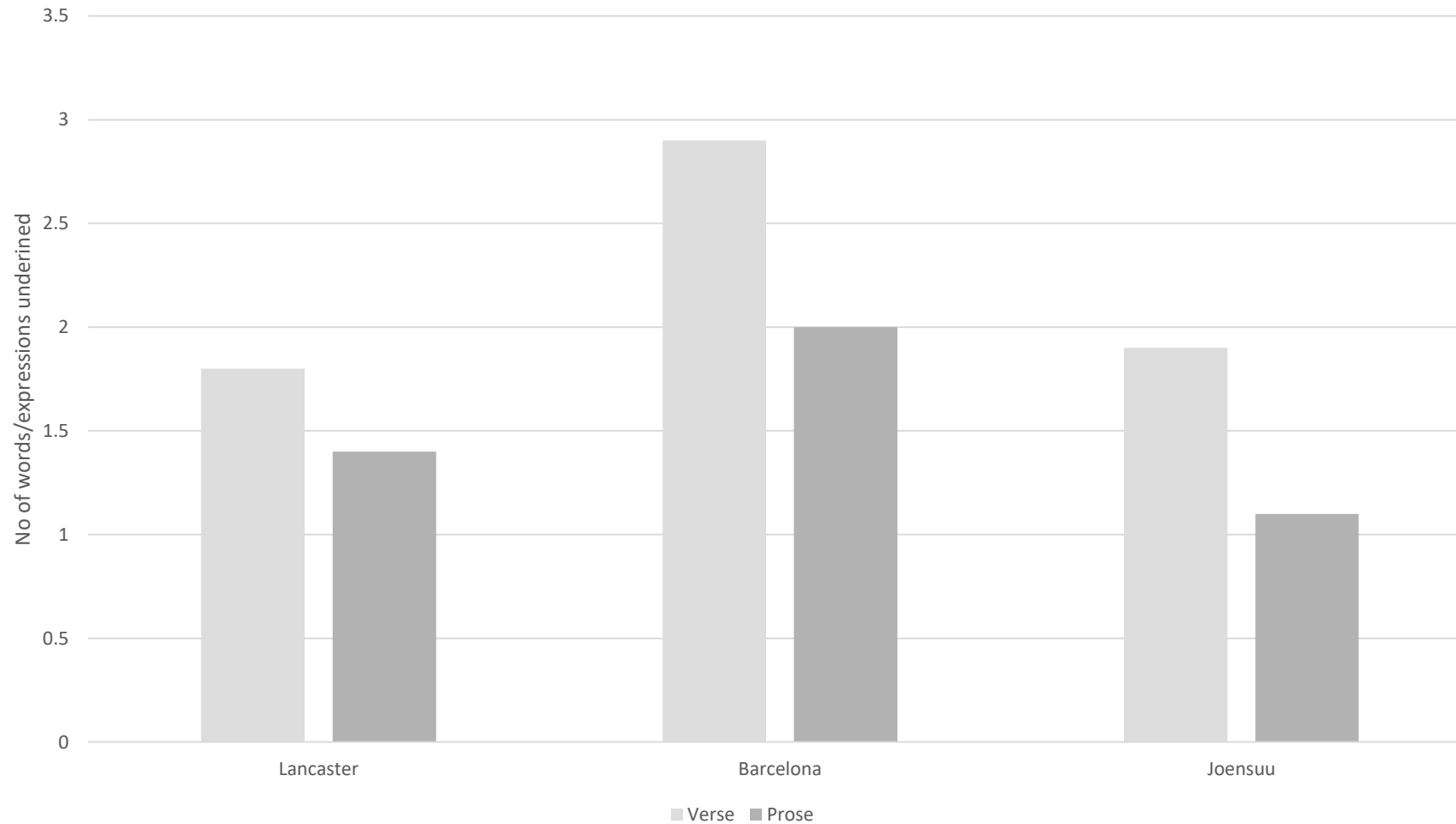
Difficulty of play extracts (Task 1)



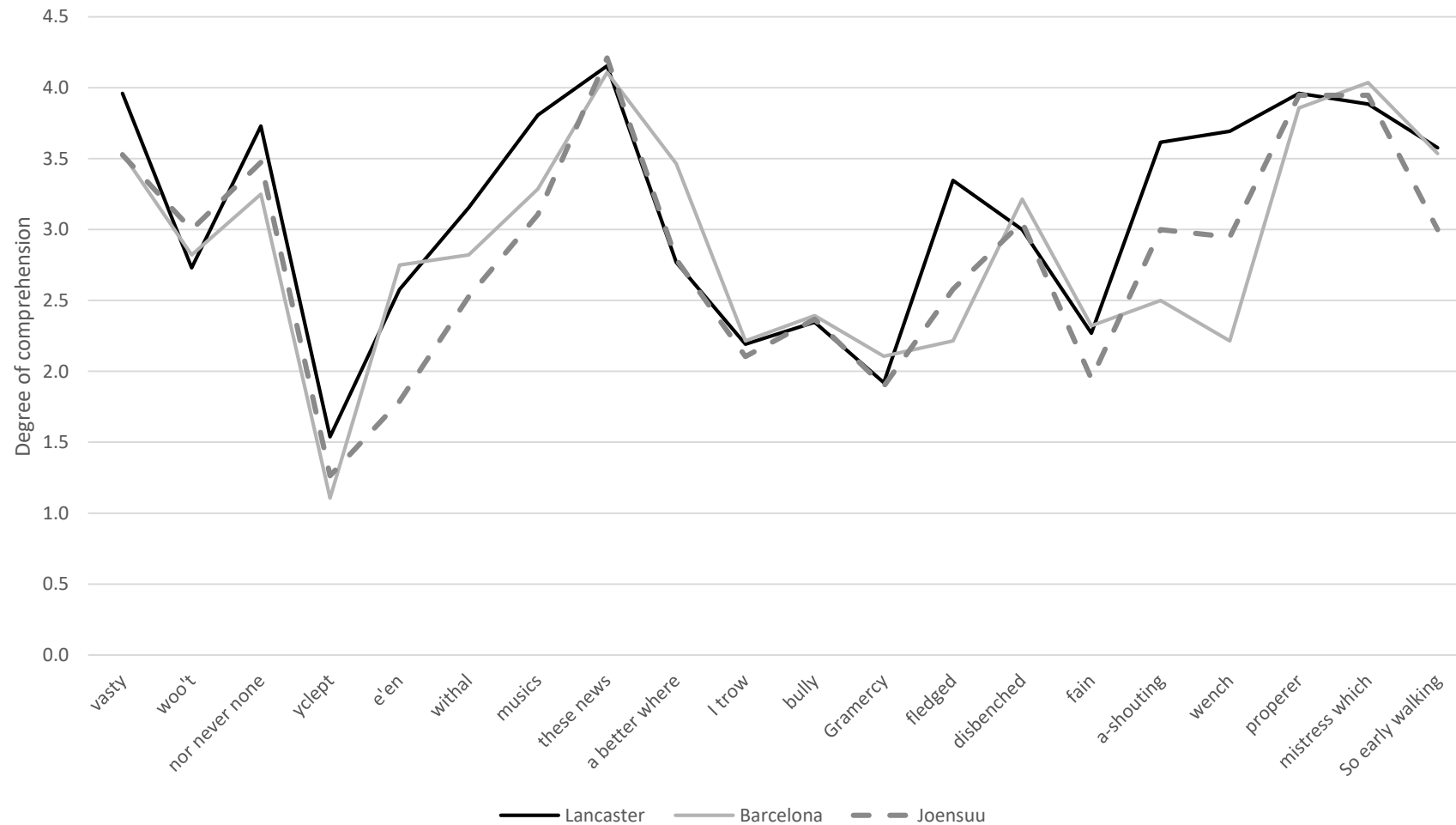
Difficulty of play extracts (Task 1)

<i>Taming of the Shrew</i>	<ul style="list-style-type: none"> • And till [my falcon] stoop she must not be full-gorged, • Another way I have to man my haggard
<i>King Lear</i>	<ul style="list-style-type: none"> • Mend your speech a little / Lest you may mar your fortunes. • Good my lord, / You have begot me, bred me, loved me
<i>Richard III</i>	<ul style="list-style-type: none"> • Was ever woman in this humour wooded? • But the plain devil and dissembling looks?
<i>Hamlet</i>	<ul style="list-style-type: none"> • If the man go to this water / and drown himself, it is, will he nill he, he goes • Argal, he that is not guilty of / his own death shortens not his own life.
<i>Much Ado About Nothing</i>	<ul style="list-style-type: none"> • You are thought here to be the most senseless and fit man for the constable of the watch • This is your charge: you shall comprehend all vagrom men
<i>Henry V</i>	<ul style="list-style-type: none"> • Tell him we could have rebuked him at Harfleur, but that we thought not good to bruise an injury till it were full ripe

Difficulty: verse / prose (Task 1)



Difficulty: words / phrases (Task 2)



What's tricky? (Task 3)

archaic
vocabulary

metaphors

word order /
sentence
structure

change in
meaning

word play

cultural
references

Summary of difficulties in rank order

Extracts	Individual items	Self-reported
<ul style="list-style-type: none"> ● use of metaphor ● lack of context ● difficult vocabulary 	<ul style="list-style-type: none"> ● archaisms ● contractions ● false friends ● coinages ● borrowings 	<ul style="list-style-type: none"> ● lexis ● semantic change ● complex syntax ● cultural references

CQPweb (Lancaster University)

ESC: First Folio Plus (1623) (<https://cqpweb.lancs.ac.uk/shakfinc006/>)

ESC: EEBO-TCP Segment (1560-1640)

(<https://cqpweb.lancs.ac.uk/shakeebotest/>)

- *fain*
- *nor never none*

would fain / fain would in Shakespeare and Early English Books Online (EEBO)

<i>would fain / fain would</i>	Shakespeare	Most frequent collocates (number of instances)	EEBO: TCP-Segment restricted query 1590-1615	Most frequent collocates (number of instances)
Relative frequency (instances per million words)	<u>33.702</u>	<i>I</i> (<u>21</u>), <i>have</i> (<u>12</u>)	<u>29.097</u>	<i>I</i> (869), <i>have</i> (751)
Typical patterns and relative frequency (instances per million words)	<i>I would fain</i> (<u>18.295</u>) <i>fain would I</i> (<u>4.815</u>)	<i>I would fain</i> <u>see</u> <i>fain would I</i> <u>go</u>	<i>I would fain</i> (<u>7.791</u>) <i>fain would I</i> (<u>2.088</u>)	<i>I would fain</i> <u>know</u> <i>fain would I</i> <u>know</u>

References

-
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Crystal, D. (2012). *Think on my words: Exploring Shakespeare's language*. Cambridge: Cambridge University Press.
- Donaldson A. (2016). *All the world's: How Shakespeare is viewed around the globe and the role his work can play to support the UK's soft power*/ed. by M. Lotten. S.n. British Council.
- Hardie, A. and Dorst, I van. (2020). A survey of grammatical variability in Early Modern English drama. *Language and Literature* 29(3): 275-301.
- Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3): 380–409.
- Johnson, K. (2014). *Shakespeare's English: a practical linguistic guide*. Harlow: Pearson.
- Leech, G.N. and Short, M.H. (1981). *Style in fiction: A linguistic introduction to English fictional prose*. London: Longman.
- Murphy, S, Culpeper, J, Gillings, M. and Pace-Sigge, M. (2020). What do students find difficult when they read Shakespeare? Problems and solutions. *Language and Literature* 29(3): 302-326.
- Murphy, S. (2019). Shakespeare and his contemporaries: Designing a genre classification scheme for Early English Books Online 1560–1640, *ICAME Journal*, 43(1), 59-82.