Statistics and visualisations of theatre corpora using corpus analysis software

Hugo Sanjurjo-González[1] and Olaia Andaluz-Pinedo[2]
[1] University of Deusto
[2] University of the Basque Country UPV/EHU

1 INTRODUCTION

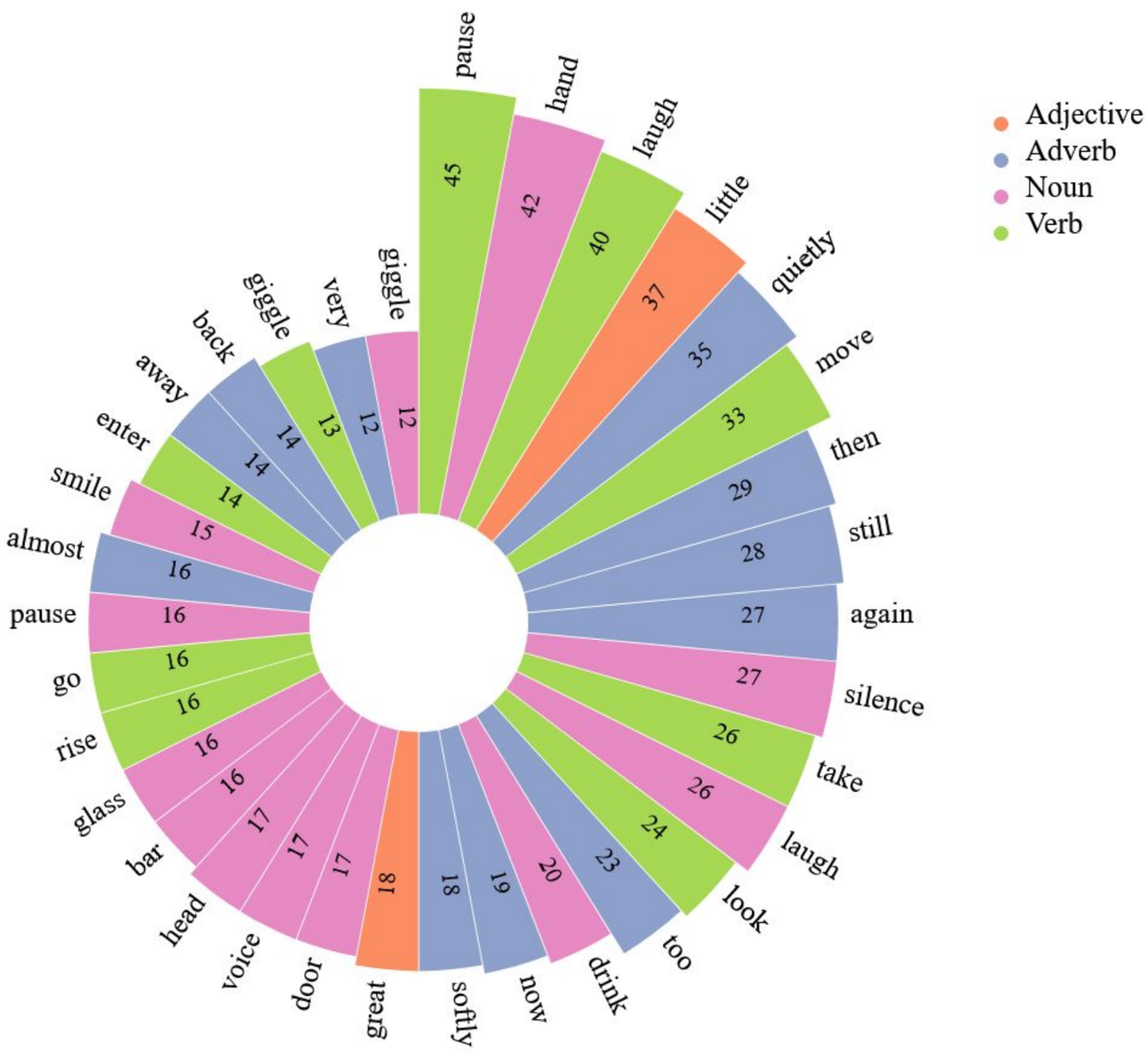
As Brezina (2018) points out, "statistics is crucial for corpus linguists because it helps us work effectively with quantitative information". Furthermore, visualisations of statistical analyses offer insights into data that may not be perceived as easily by reading numbers. However, statistical techniques in available corpus software do not take into account the peculiarities of theatre corpora. Play-texts have a specific structure that includes acts, scenes, utterances, speakers, dialogues and stage directions, and these parts often need to be operated on for analysis.

2 AIMS

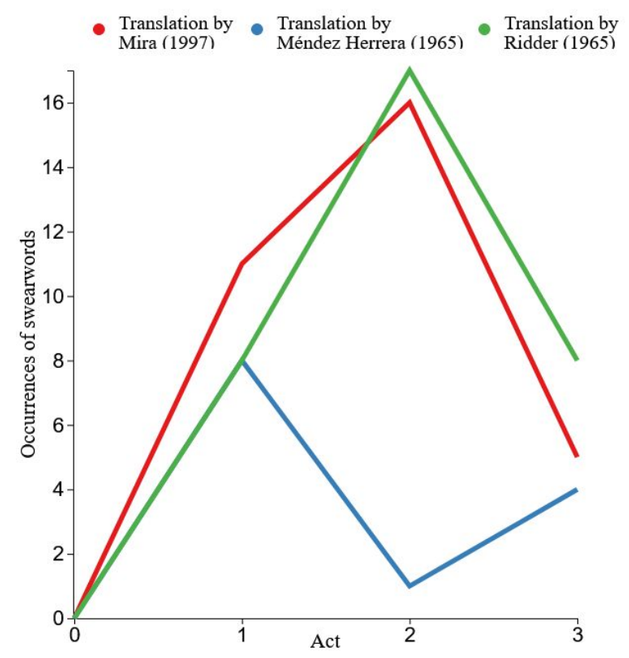
In this poster, we suggest some examples of statistics and visualisations that apply and interrelate theatre-specific filters, based on the units of acts, speakers, utterances, stage directions and dialogues, instead of using play-texts as wholes. These analysis options will be included in the corpus tool ACTRES Corpus Manager (Sanjurjo-González 2017), allowing for a novel approach that opens up further possibilities for the study of theatre corpora.

3 PRELIMINARY RESULTS

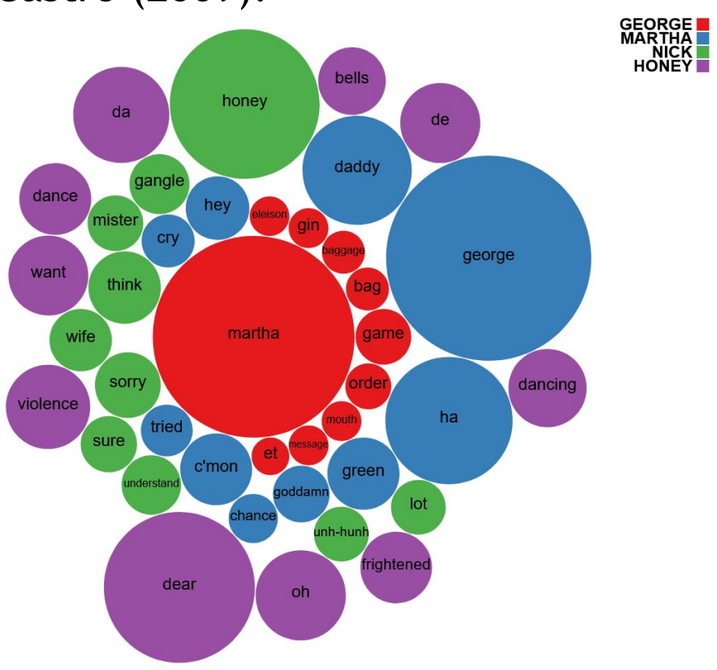
Visualisations based on Edward Albee's play *Who's Afraid of Virginia Woolf?* are included below in order to illustrate statistical analyses developed that use theatre structural filters.



The circular barplot shows the most frequent lemmas found in stage directions and their POS tags at a glance.



The number of swearwords in each act of three translations are compared, following Gómez-Castro (2009).



The keywords of each character's dialogue in the play are displayed in this bubble chart.

4 DISCUSSION

The usefulness of statistics and visualisations for corpus linguistics has been highlighted in the literature (Brezina 2018). The application of these techniques to theatre texts is still not integrated in corpus software. However, the work-in-progress presented suggests that theatre-specific visualisations are a helpful way of displaying complex analyses that include different types of annotation and theatre structural mark-up. These advances will enhance analysis possibilities of monolingual, comparable or parallel theatre corpora.

REFERENCES

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

Gómez-Castro, C. (2009). *Traducción y censura de textos narrativos inglés-español en la España franquista y de transición: TRACEni (1970-1978)* (Doctoral dissertation). University of León, León, Spain.

Sanjurjo-González, H. (2017). *Creación de un corpus para el análisis de corpus lingüísticos* (Doctoral dissertation). University of León, León, Spain.

Deusto
Universidad de Deusto
University of Deusto

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

AESLA 38