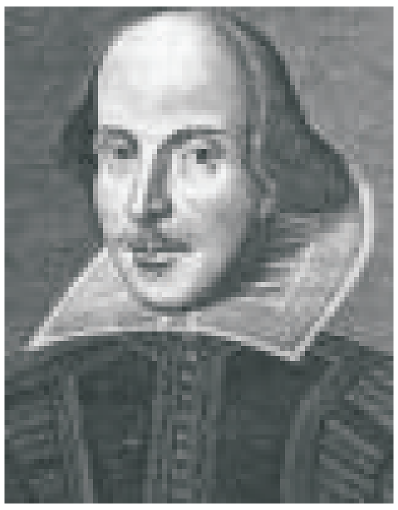
Computational Stylistics Facility

1. Purpose

The CSF aims to make some key techniques of computational stylistics widely available. In particular, it allows any user to produce a Principal Component Analysis to flexible parameters in minutes. It is designed to be so quick and simple in operation that in a single session the user can test a hypothesis about the language of Shakespeare, use the results to create a new one, and test that in turn.



2. Capabilities

The texts can be analysed by play, by blocks within plays, and by characters. The plays and characters are tagged so that subsets of them can be readily created: the plays by genre, date and so on, and the characters by size of part, by gender, and by social category. Users create word lists to provide the variables for the PCAs within the CSF, by choosing the most common words of the whole corpus, or a sub-corpus, or by pasting or typing in their own list.

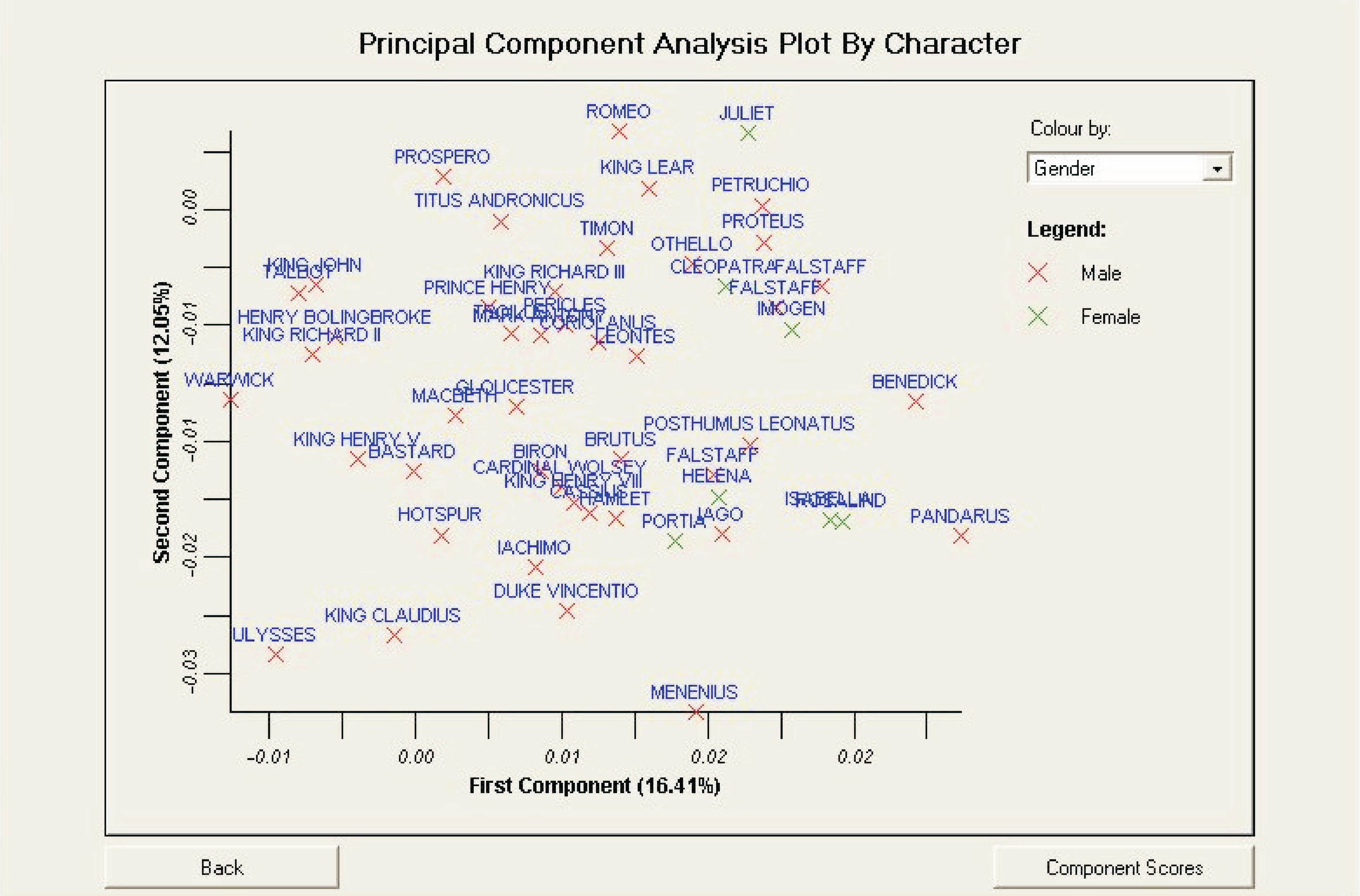
The CSF will display frequencies of any variable, listed according to the segmentation chosen, and users can inspect individual instances of the word-variables in context. The CSF incorporates a layer which expands contractions in the base text (the public-domain Moby Shakespeare). Instances of *I* and *will* contained in the contraction *I'll* are counted automatically, for example.

3. History

In 2002 the ARC funded the Australian e-Humanities Network through its Learned Academies Special Projects scheme. The Centre for Literary and Linguistic Computing at Newcastle, one of the partners in the Network, undertook to create a Computational Stylistics Facility to further the diffusion of computational methods in the humanities.

4. Case Study

A typical starting point would be the question, do Shakespeare's women characters speak differently from his men, and if so, in what way? To explore this we might select for analysis all characters who speak more than 3,000 words of dialogue, and use the 50 most frequent words in the Shakespeare corpus overall. (For other purposes, one could choose just pronouns to count, or military terms, or agricultural ones, following any number of particular interests.) The PCA analysis for the study of the larger characters looks like this:

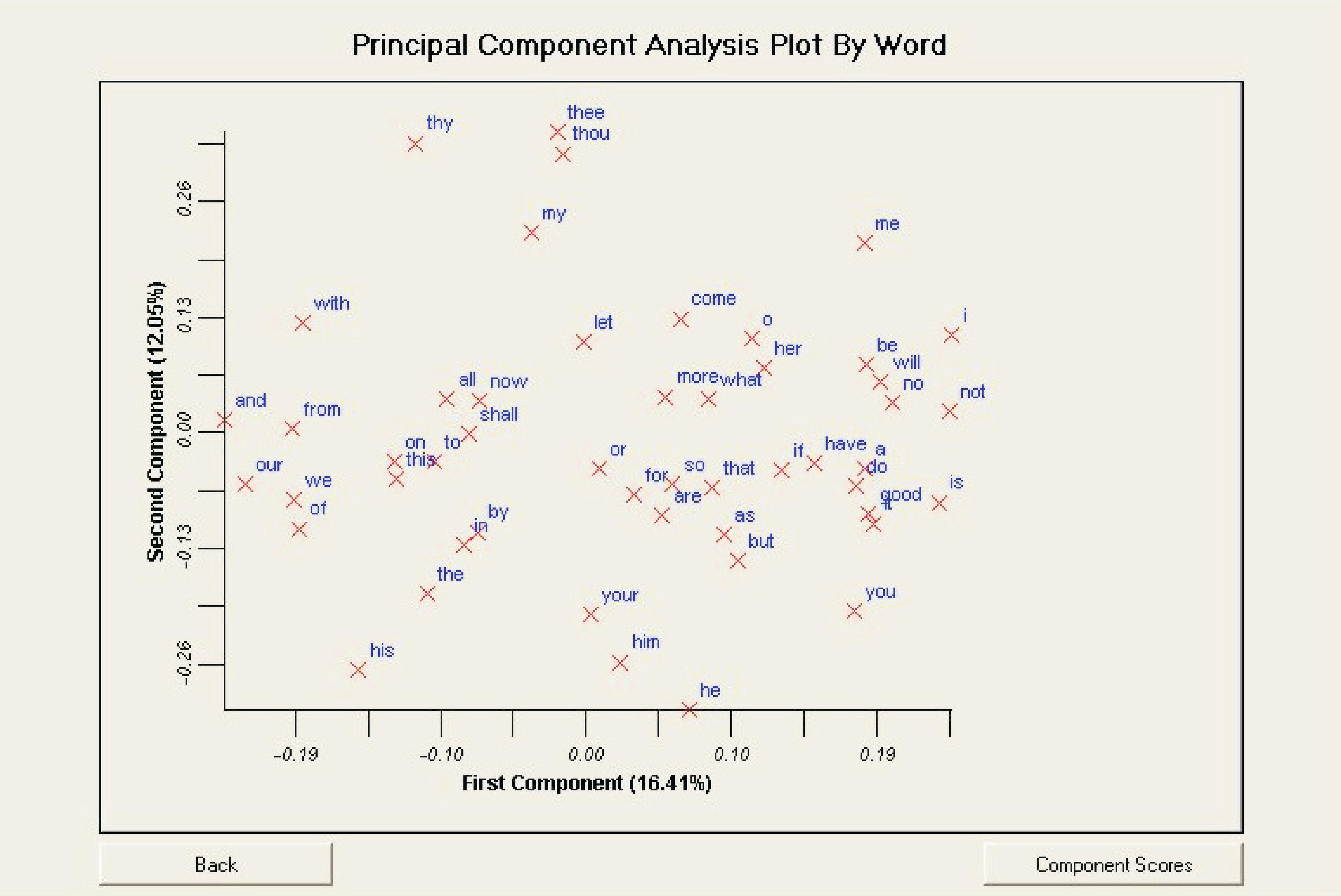


Principal Component Analysis Plot By Character

Colour by: Gender

The fifty characters with their labels make for a crowded graph, but the coordinates of individual entries can be displayed on a separate screen for identification. The two women characters to the lower right are Rosalind and Isabella.

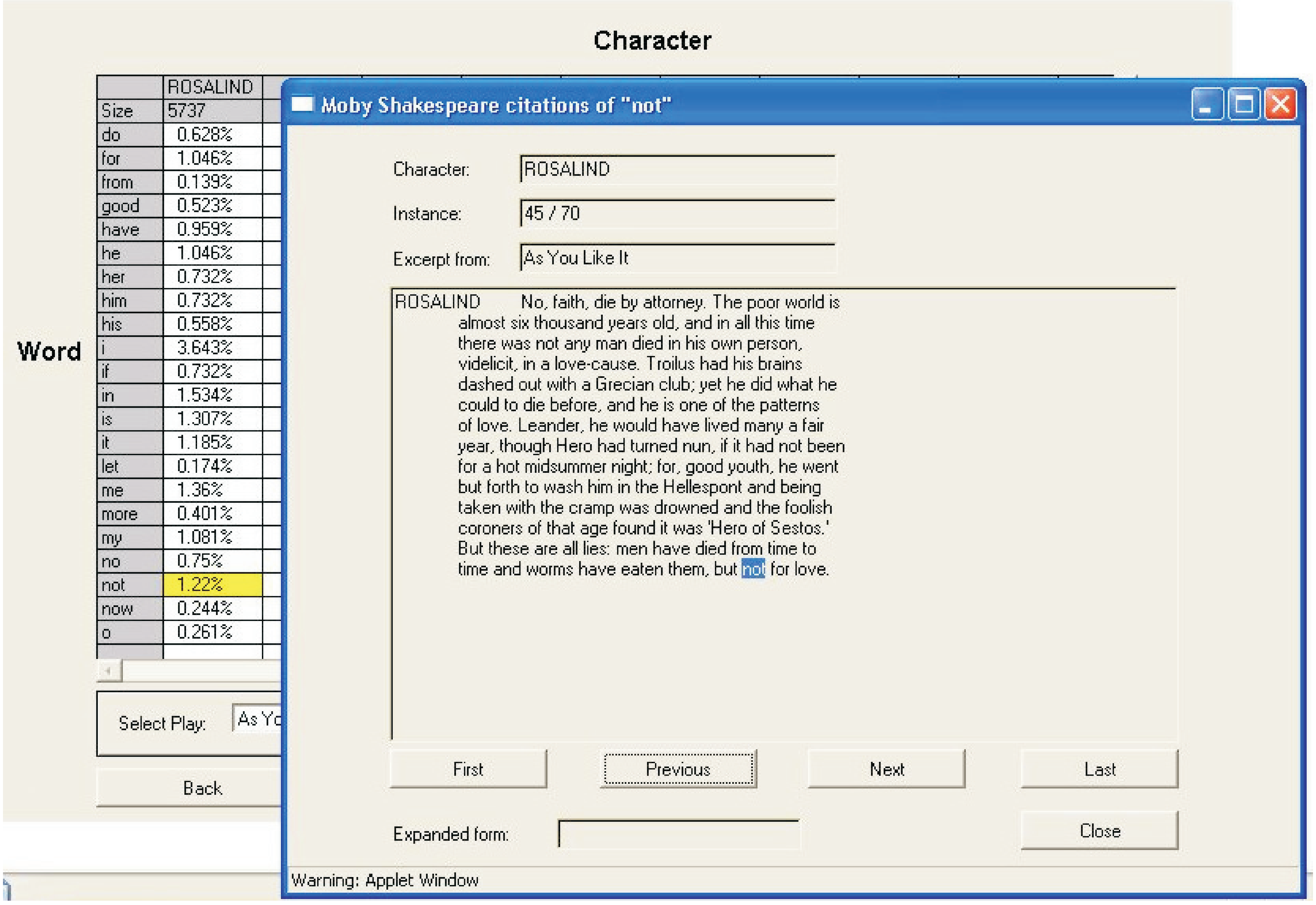
The seven women characters do behave to a degree as a group, being all to the higher-value end of the horizontal axis. They are not the outliers. None are as far to this end as Benedick from *Much Ado about Nothing* or Pandarus from *Troilus and Cressida*. We can now compare the loadings for the variables on the same principal components.



Principal Component Analysis Plot By Word

The position of the words and the characters correspond. A word with a high value on the First Component is likely to occur unusually often in a character with a high value on the same component. Words which appear together behave alike; they tend to be abundant, or scarce, in the same texts. It looks as though the First Component is an axis of differentiation between styles of pronouncement and public import (*and*, *our*, *with*, *from*) and those of personal, interactive dialogue (*I*, *not*, *is*, *no*). On the characters graph, correspondingly, we find kings and nobles from history plays at the public pronouncement end, and bantering characters from comedies and satirical plays at the other.

Looking more closely, we can seek the percentage frequency of some of these highlighted words in these same characters, and look at actual occurrences. Here, for instance, is data for *not* in the spoken dialogue of Rosalind, with one of the instances of her use of the word.



	ROSALIND
Size	5737
do	0.628%
for	1.046%
from	0.139%
good	0.523%
have	0.959%
he	1.046%
her	0.732%
him	0.732%
his	0.558%
i	3.643%
if	0.732%
in	1.534%
is	1.307%
it	1.185%
let	0.174%
me	1.36%
more	0.401%
my	1.081%
no	0.75%
not	1.22%
now	0.244%
o	0.261%

Select Play: As Yo

Character: ROSALIND

Instance: 45 / 70

Excerpt from: As You Like It

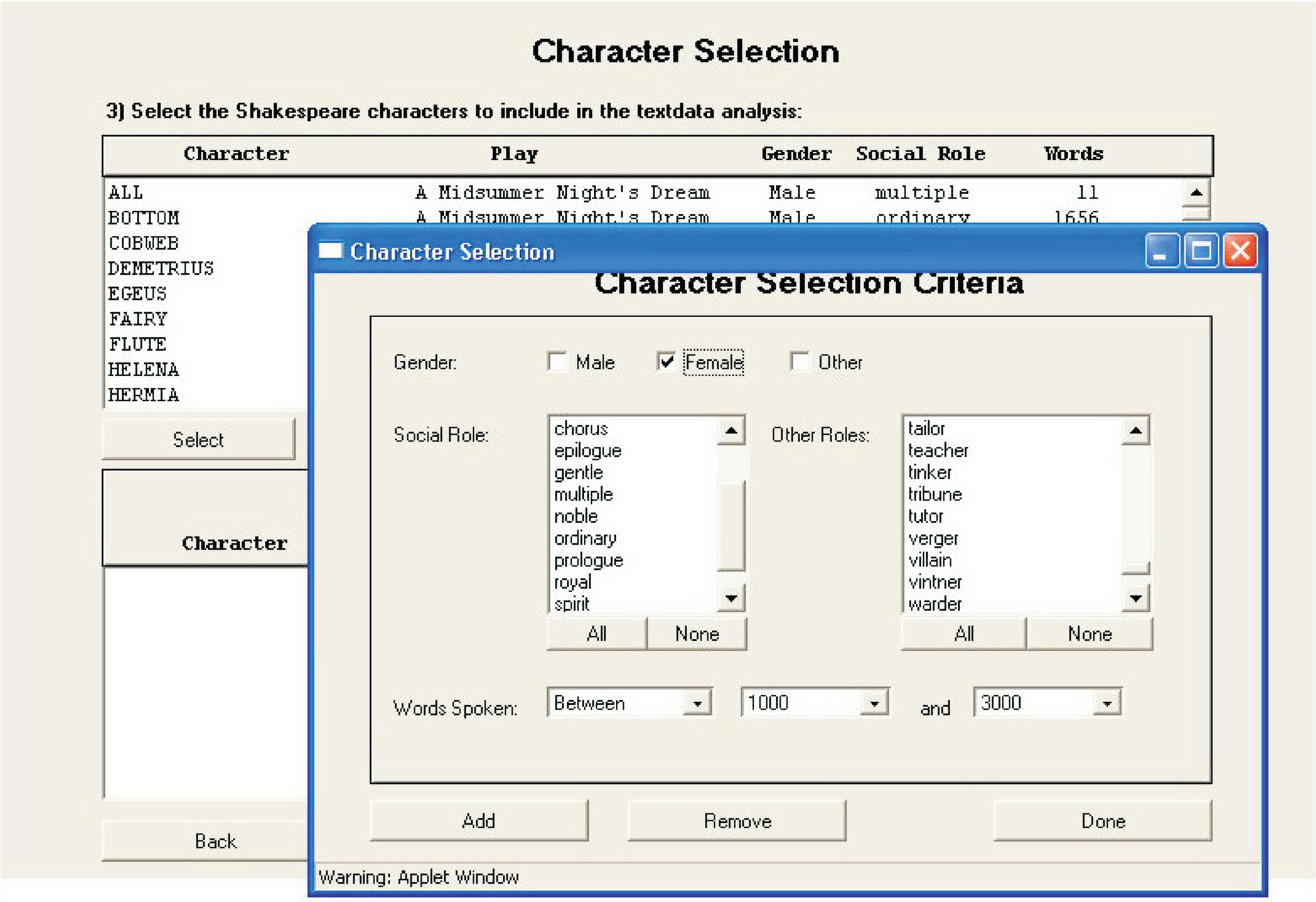
ROSALIND No, faith, die by attorney. The poor world is almost six thousand years old, and in all this time there was not any man died in his own person, videlicit, in a love-cause. Troilus had his brains dashed out with a Grecian club; yet he did what he could to die before, and he is one of the patterns of love. Leander, he would have lived many a fair year, though Hero had turned nun, if it had not been for a hot midsummer night; for, good youth, he went but forth to wash him in the Hellespont and being taken with the cramp was drowned and the foolish coroners of that age found it was 'Hero of Sestos.' But these are all lies: men have died from time to time and worms have eaten them, but not for love.

Expanded form:

Women characters are more interlocutory than most of the men, it seems, and less given to general pronouncements. In this sort of contrast there are other, male characters who are yet more interlocutory.

5. Following on

To take this further one might go in any number of different directions within the CSF. What happens if women with smaller spoken parts are included as well? What happens if the analysis is confined to comedy, leaving out the grandly spoken kings and nobles? Why do women in Shakespeare say *not* and *no* so much? Then there are all the other questions of genre and chronology to explore, and all the groupings suggested by the character criteria screen below. The CSF awaits your questions.



Character Selection

3) Select the Shakespeare characters to include in the textdata analysis:

Character	Play	Gender	Social Role	Words
ALL	A Midsummer Night's Dream	Male	multiple	11
BOTTOM	A Midsummer Night's Dream	Male	ordinary	1656
COBWEB				
DEMETRIUS				
EGEUS				
FAIRY				
FLUTE				
HELENA				
HERMIA				

Character Selection Criteria

Gender: □ Male ☑ Female □ Other

Social Role: chorus, epilogue, gentle, multiple, noble, ordinary, prologue, royal, spirit

Other Roles: tailor, teacher, tinker, tribune, tutor, verger, villain, vintner, warder

Words Spoken: Between 1000 and 3000

Warning: Applet Window

The CSF system is a distributed Java based application running on a UNIX server. It comprises a server side component that processes requests and a web-based Java applet that provides a graphical user interface for users to supply a flexible range of input parameters to the server and in turn interpret the results in various forms. The system was designed and written by Russell Whipp in Newcastle following a detailed brief from the CLLC. After testing it will be made available through the Australian e-Humanities Gateway at www.ehum.edu.au.

Hugh Craig
Centre for Literary and Linguistic Computing,
University of Newcastle