Conversation, characterisation and corpus linguistics: Dialogue in Jane Austen's *Sense and Sensibility*

E.H. Hubbard
Department of Linguistics
University of South Africa
PRETORIA
E-mail: hubbaeh@unisa.ac.za

Abstract

Conversation, characterisation and corpus linguistics: Dialogue in Jane Austen's *Sense and Sensibility*

*This article reports on a corpus-based exploration of the role that fictional dialogue plays in characterisation. The focus is on the two main characters of Austen's **Sense and Sensibility** and (a) the extent to which certain features of their dialogue can be said to tie in with general perceptions that Elinor represents the "sense" and Marianne the "sensibility" of the novel's title; and (b) the extent to which Austen can be said to have exploited these features to enable the sisters to speak with subtly differing voices. The features themselves were drawn from two linguistic frameworks, namely cohesion in text linguistics (specifically, the category of conjunctive cohesion as originated by Halliday and Hasan (1976)), and the category of "involvement" in register analysis (most prominently, Biber 1988). The density of these features in each dialogue was calculated, compared statistically and salient differences considered in relation to the focal issues of the study. Although two of the five hypotheses formulated were not supported, the results overall provided strong indications that Austen successfully distinguishes between the sisters through their dialogue, and often in ways that link with less subtle, more explicit cues to their character that are given in the text. The study thus reveals how certain text-linguistic and register features can underpin characterisation in fiction, and in so doing explicates aspects of what it is that readers and literary critics respond to when they comment on characterisation in a novel.*

> Elinor, this eldest daughter whose advice was so effectual, possessed a strength of understanding, and a coolness of judgement ... She had an excellent heart; her disposition was affectionate, and her feelings were strong; but she knew how to govern them ...

> Marianne's abilities were, in many respects, quite equal to Elinor's. She was sensible and clever, but eager in every thing; her sorrows, her joys, could have no moderation. She was generous, amiable, interesting; she was every thing but prudent (Austen, 1959:30).

1. "Sense" and "sensibility"

It is by way of the above unabashed, direct descriptions that Jane Austen introduces us to Elinor and Marianne, whom readers cannot help linking with the "sense" and the "sensibility", respectively, of the novel's title. Indeed, as Williams (1986) reports, some critics have seen the strength of this association as a serious flaw in Austen's first published novel, claiming that:

> … there is an unyielding schematic antithesis, declared in the title and made too explicit in the persons of the two central characters, and that the novel always demonstrates the superiority of Elinor's Augustan sense over Marianne's Romantic sensibility, so that it does not really make use of the dynamics inherent in an antithesis (Williams, 1986:31).

However, as is evident even in her opening description of the sisters, Austen herself gives notice that we should not see them simply as opposites. Thus although she knows how to govern them, Elinor's feelings are strong, she has an excellent heart and affectionate disposition, while Marianne is in many ways Elinor's equal and is clever.

Marianne is, it is true, also "sensible", and in Austen's time this word, like the "sensibility" of the title, did not normally "connote that hard-headed, cool, practical type of 'common sense' it usually suggests today" (Stokes, 1991:129). Although this meaning had been available at least since the 16th century, it appears that until modern times it was rarely applied and not regarded as appropriate usage: "Stigmatized by Johnson as used only 'in low conversation'" (CEOED, 1971: s.v. "sensible").

Marianne's "sensibility" relates rather to a capability for feeling (Stokes, 1991:129), the "power or faculty of feeling, capacity of sensation and emotion as distinguished from cognition and will" (CEOED, 1971: s.v. "sensibility"): hence a meaning close to modern "sensitivity" and quite distinct from "sense", which in Austen's work is essentially synonymous with prudence and (good) judgement (Stokes, 1991:126).

The initial descriptions of the sisters, then, are nuanced rather than merely antithetical, and we should not expect, in an Austen novel, that the delineation of the two main characters would descend into caricature. All the same, it is easy enough to agree with Gilbert Ryle's claim that "Sense and Sensibility really is about the relations between Sense and Sensibility or, as we might put it, between Head and Heart, Thought and Feeling, Judgement and Emotion" (quoted in ApRoberts, 1986:46). In focusing on the relationship between these elements, Ryle does not deny that there may be times when Elinor reveals sensibility and Marianne sense, or that other characters too might be assessed in terms of the opposition, but the thrust of his argument is that the two sisters are its chief exponents.

This, notwithstanding the nuances in her opening description, is clearly the view of Austen herself. How, then, does she sustain this differentiation of the sisters throughout the rest of the novel? The present paper aims to provide a partial answer to this question by applying a corpus-based approach, informed by text linguistics and register analysis, to the main reflection of the sisters' linguistic behaviour: their conversation.

2. Characterisation and conversation

Despite the importance of characters in literature, various writers have commented on the scarcity of work that theorises character (e.g. Chatman, 1972; Van Peer, 1989), though Culpeper (2001:1) notes that there have at least been two recent high profile journal issues devoted specifically to characterisation (*Poetics Today,* 1986 and *Style,* 1990). A reading of these special editions, however, reveals that none of their articles provide or suggest a systematic linguistic account of features of characters' dialogue as a way either to a better understanding of readers' responses to them or to a fuller appreciation of the craft of the writer.

Culpeper (2001) is a substantial attempt at addressing the dearth of literature on characterisation, particularly with respect to drama. In his discussion of textual cues in characterisation he distinguishes authorial cues, "where character information comes relatively directly from the author, as in the case of stage directions", explicit cues, "where we find

characters explicitly presenting themselves or others" and implicit cues, "where we have to infer [...] character information from linguistic behaviour" (Culpeper, 2001:164).

The author or narrator in fiction has a lot more scope for intervening directly in the portrayal of characters than the dramatist, who is limited essentially to stage directions. Thus Culpeper's categorisation needs to be modified for application to the novel, allowing for two kinds of authorial cue to match the non-authorial ones – explicit and implicit.

We could identify among explicit authorial cues the author's or the narrator's direct comments on or descriptions of characters (as when Austen introduces Elinor and Marianne). Explicit non-authorial cues are then instances where characters comment specifically about themselves or are commented on by other characters.

Implicit authorial cues are found where readers are not informed directly about characters' personalities but have to make inferences about them based on descriptions of the actions, thoughts and speech of the characters. Implicit non-authorial cues, the category most relevant to the concerns of the present paper, are found when the thoughts and speech are not described by the author or narrator but are represented as emanating directly from the characters themselves.

The relationship between implicit authorial and implicit non-authorial cues is not based on a simple distinction. There is a variety of ways in which speech (and in essentially parallel fashion, thought) can be represented, and Short (1996:293), for example, refers to a gradation or cline in terms of the degree of apparent narrator control over what is reported. Moving from the authorial to the non-authorial end of the cline (and with examples made relative to one actual narrator statement from *Sense and Sensibility*) are the categories: narrator's representation of speech (*Elinor talked*); narrator's representation of speech acts (*Elinor was again obliged to decline her invitation* (Austen 1959:135)); indirect speech *(Elinor said that she would not be able to go to Cleveland)*; free indirect speech (*She would not be able to go to Cleveland*); and direct speech (*I am afraid that I cannot go to Cleveland*).

It is, then, in the presence of direct speech that the mediation of the author or narrator is least visible and where, arguably, the reader experiences the highest degree of immediacy with respect to the relevant character. On the basis of the implicit cues provided in direct speech (thus, through dialogue), the reader develops an understanding of a character in a manner that could be said to more closely resemble this process in real life, as compared to the way the process works when the

cues are authorial. It might also be argued that in the creation of a novel the drawing of character through dialogue poses special challenges, not least because the *cues* need to be implicit and indirect. It is one thing for the authorial voice to say that Elinor knew how to govern her feelings or that Marianne was eager in everything; quite another to convincingly represent such characteristics through subtle differences in dialogue.

Jane Austen has been lauded as "one of the greatest, because one of the most accurate, writers of dialogue of her own or any age" (Chapman, in Burrows, 1987:108). The present article tests this view with reference specifically to the dialogue of Elinor and Marianne. The focus is restricted to dialogue, that is, direct speech, and excludes the related implicit, non-authorial cues generated by direct thought representation. This focus can be justified because there is a great deal more dialogue in *Sense and Sensibility* than direct thought, with the proportion of text devoted to the latter in this novel being, at less than five percent, the lowest in all Jane Austen's works (Burrows, 1987:166). There is also a logistical justification in that, unlike many modern writers, Austen is extremely consistent in marking off her dialogue with inverted commas. This made the compilation of dialogue corpora for Elinor and Marianne a relatively simple task.

3. Corpus linguistics and characterisation

Although Short *et al.* (1996:112) lamented the lack of scholars interested in both corpus linguistics and stylistics, this picture is now changing quite rapidly. Even in the more specific field of corpus linguistics and the stylistics of literary texts, the last few years have for instance seen the emergence of a number of local or locally published studies (e.g. Watson, 1999, Zyngier, 1999 and Kruger, 2000).

Very few studies have to date used a corpus approach to link dialogue to characterisation, but the major exception is Burrows (1987). He makes the point that in most discussions of English fiction a large proportion of the language is simply ignored, namely the most common 20, 30 or 50 words in the relevant work, which in Jane Austen's oeuvre make up respectively about a third, two-fifths, and a half of the running words of her texts. He goes on to argue that:

> The neglected third, two-fifths or half of our material has light of its own to shed on the meaning of one novel or another; on subtle relationships between narrative and dialogue, character and character; on less direct and less limited comparisons between novels and between novelists; and ultimately on the very process of reading itself (Burrows, 1987:1).

The focus on commonest words is typical of a fairly long tradition of scholarship aimed at the resolution of authorship disputes (such as whether Bacon or Marlowe, for instance, really wrote some of the plays normally attributed to Shakespeare: Morton (1978) provides an overview of this field). One key assumption about these studies is that, as the commonest words are mostly function words rather than full lexical items, their patterns of use are reflexive rather than reflective: they represent stylistic traits that are instinctive; choices that are largely subconscious and thus less likely to change across different texts written by the same author. One of Burrows' findings is that the relative frequencies of some of these words in Jane Austen's writing do differ significantly from those of certain other writers, but a conclusive assessment does not emerge.

What, though, of the language of different characters in novels? Could authors, despite the constraints of their supposedly "built-in" patterns of use of these common, mostly non-lexical function words, nevertheless create characters who in their dialogue would reveal their own separate and consistent patterns? Burrows' main research findings indicate that, for Jane Austen at any rate, the answer to this question is a resounding "yes". So much so, that he suggests that novels would not be reliable material for authorship attribution if authors in general had "Jane Austen's capacity to 'change style' from character to character" (Burrows, 1987:99).

Burrows' positive findings on dialogue and character differentiation in Austen's work as a whole were one stimulus for the present article. Burrows (1987) does also make occasional references to the use of a few of the commonly occurring words in the dialogue of Elinor and Marianne, but his work differs from this article not only in being much more synoptic but also in that it is strongly data driven: its point of departure is the most frequent 20 to 50 words in Austen's oeuvre and how their distribution profiles differ, mostly in relation to dialogue. On the basis of statistical findings on certain words and groups of words he then induces generalisations about the role of these words in characterisation.

The present article, on the other hand, in seeking to explore how and to what extent the sense versus sensibility distinction gets to be woven into the dialogue of Elinor and Marianne, is a more specific study. It also takes a more theory-driven, hypothetico-deductive approach, the two frameworks that inform it being drawn from text linguistics and register analysis.

4. Analytical frameworks, hypotheses and procedures

What kind of linguistic frameworks might throw light on the essentially literary-theoretical issue of the role that dialogue plays in the development of fictional characters? More specifically, which frameworks might be relevant to the elicitation and analysis of linguistic features in dialogue that associate with the opposing character traits (as noted by Ryle) of sense and sensibility, judgement and emotion, thought and feeling, and head and heart?

Two frameworks were selected as being potentially appropriate to this purpose, namely those offered by the text linguistic category of conjunctive cohesion (originating with Halliday and Hasan, 1976) and the register category of "involvement" (most saliently, Biber, 1988).

Conjunctive cohesion links the meanings of clauses, sentences or other textual units as wholes, expressing "the way in which what is to follow is systematically connected to what has gone before" (Halliday & Hasan, 1976:227). The main subcategories of conjunctive cohesion are:

- **additive conjunction**, which connects textual units essentially in terms of an "and" or "or" relation, as exemplified also by conjunctive expressions such as *furthermore* and *in other words*;

- **adversative conjunction**, which connects units basically in terms of a "contrary to expectation" relation (Halliday & Hasan, 1976:250), exemplified by conjunctives such as *but, however* and *although*;

- **causal conjunction**, which connects units in terms of causal and conditional relations, exemplified by conjunctives such as *so, because, therefore, for this reason* and *if*; and

- **temporal conjunction**, which connects units in terms of time relations, exemplified by conjunctives such as *next, while* and *by this time*.

The assumption is made that causal relations will tend to be associated more strongly with the discourse of rationality, of judgement, of "sense", than are additive and temporal relations. Similarly, expressions such as adversatives relate to "cognitive acts that make discriminations – the processes of distinguishing, making exceptions, conceding or contrasting by which thinking, and the prose which represents thinking, is carried on" (Fahnestock, 1983:415). Given these assumptions, and the secondary assumption that discourse with relatively more of such relations will tend to contain relatively more of the explicit signals of such relations – the conjunctives – the first two hypotheses of this study can be formulated as directional hypotheses, as follows:

- **H1: Density of adversative conjunction:** In *Sense and Sensibility*, Elinor's dialogue will reveal a higher density of adversative conjunctives than that of Marianne.

- **H2: Density of causal conjunction:** In *Sense and Sensibility*, Elinor's dialogue will reveal a higher density of causal conjunctives than that of Marianne.

  No matter what the findings on these two directional hypotheses might be, a third general hypothesis tests, without predicting directions, whether the sisters' patterns of use of conjunctive cohesion in general are significantly different from one another.

- **H3: Patterns of conjunction:** In *Sense and Sensibility*, the dialogue of Elinor and of Marianne will differ in terms of the overall relative density of use of the four main types of conjunction.

The second set of hypotheses in the study derives from Biber's multidimensional approach to register variation and, most pertinently, the dimension of *involvement* within his framework (Biber, 1988). This term was first given wide exposure by Tannen, who contrasted linguistic features that signalled speaker or writer focus on interaction and interpersonal involvement – typical of spontaneous, relatively unplanned discourse – with features that signalled a focus on communicating information – typical of more closely planned discourse (e.g. Tannen, 1985). This opposition has since been validated by Biber, who was able to say, on the basis of statistical and functional analysis of a substantial corpus, that it is "an extremely powerful factor representing a very basic dimension of variation among spoken and written texts in English" (Biber, 1988:104).

How might this dimension of variation show up differences between the dialogue of Elinor and Marianne? The initial descriptions of the sisters indicate that Elinor possesses a coolness of judgement and is able to govern her feelings; Marianne, on the other hand lacks her sister's prudence and moderation, and is "eager in everything". It is then reasonable to assume that if these characteristics are carried over into the sisters' dialogue, Marianne's will tend to be less planned, more spontaneous and more focused on interpersonal contact and involvement, and so Biber's most basic dimension of register variation provides the rationale for the third directional hypothesis of this study:

- **H4: Density of involvement:** In *Sense and Sensibility*, Marianne's dialogue will reveal a higher density of involvement features than that of Elinor.

Given that we are dealing with dialogue, and given that involvement was found by Biber (1988:165) to be the dimension which is most typical of all face-to-face conversation, the features of this dimension represent a very important set of parameters for comparing particular speakers' dialogue. Thus, as in the case of the conjunction hypotheses, no matter what the findings on this directional hypothesis might be, a further general hypothesis tests whether the sisters' patterns of use of involvement features in general are significantly different from one another:

- **H5: Patterns of involvement:** In *Sense and Sensibility*, the dialogue of Elinor and of Marianne will differ in terms of the overall relative density of use of the main types of involvement feature.

The database for this study was all the dialogue spoken by the two sisters, marked off consistently by inverted commas, and totalling 8905 words for Elinor and 6104 words for Marianne. The two corpora were analysed with the assistance of the Longman Mini Concordancer. With respect to the cohesion hypothesis, all instances of conjunctive expressions linking clauses or larger textual units were analysed. In testing the involvement hypothesis, some of the 23 features that Biber (1988:89) found to associate strongly with involvement, such as "THAT deletion" and "present tense verbs" could not be included for analysis because the corpora were not tagged, and the category of contractions (e.g. *can't, doesn't*) was not applicable because they are not used in this novel. However, 15 features, including eight of the ten most strongly characteristic of involvement, were analysed.

In testing the hypotheses, the Chi^2 statistic was applied to the totals for each feature (making due allowance for the difference in the sizes of the two corpora) and where one-way analyses were needed, the Yates correction factor was used (e.g. Hatch & Farhady, 1982:165-172). The significance level was set at $p \leq 0,05$, with results of $p \leq 0,01$ being labelled "highly significant" and of $p \leq 0,001$ "very highly significant". To provide some indication of what features did not differ significantly between the two sets of dialogue but nevertheless showed some statistically supportable degree of difference, cases of $p \leq 0,10$ were labelled as having "a tendency toward difference" and cases of $p \leq 0,25$ as having "a slight tendency toward difference". The $p \leq 0,25$ level also served as a requirement for the inclusion of individual expressions in Table 2 (the involvement features) so that this table could provide a focus on only those expressions that showed some minimal degree of difference in density between the two dialogues (because there were far fewer conjunctive expressions, all of these could be included in Table 1).

Similarly, discussion of findings will be concentrated almost exclusively on such expressions.

5. Findings

Results and discussion of results are dealt with together in this section, taking each hypothesis in turn.

- **H1: Density of adversative conjunction**

The density of adversative conjunction in Elinor's dialogue is not significantly higher than in Marianne's and so this hypothesis was not supported. In fact there is even a slight tendency toward difference in the opposite direction, with the density higher in Marianne's dialogue (13,6 per 1000 words, as against 11,3 for Elinor: see Table 1). This is due mostly to a slight tendency toward higher use of the most frequent adversative expression, *but*, and also of *yet*, while Elinor shows a slight tendency for preferring *however*. These marginal differences in word choice seem to be appropriate to the two characters, given the more formal, less spontaneous ring of *however* relative to *but* and *yet*, but otherwise the characters' overall use of adversatives is not distinctive.

- **H2: Density of causal conjunction**

The density of causal conjunction in Elinor's dialogue is significantly higher than in Marianne's (p≤0,05, one-tailed test: cf. e.g. Ferguson, 1981:215), and so this hypothesis is supported. The main factor here is Elinor's significantly more frequent use of *for* (in the sense of "because"), though she also registers non-significant higher values for *if* and for all the low frequency causal conjunctives except *because* (see Table 1). This finding is consonant with perceptions of Elinor's "strength of understanding" and "coolness of judgement" and is one indication of Austen's finely tuned differentiation of character through dialogue.

- **H3: Patterns of conjunction**

This hypothesis suggests that the density of the four main types of conjunction – additive, adversative, causal and temporal – relative to one another will differ between the two sets of dialogue. This was indeed found to be the case, the Chi^2 test revealing a difference significant at the p≤0,05 level. Thus although the overall proportion of conjunctions in the dialogues is very similar (34,8 per 1000 words for Elinor and 37,0 for Marianne), they reveal differential proportions of use of the four main types. This is partly, as we have seen, due to Elinor's significantly higher use of causal conjunctives and to Marianne's non-significant slight tendency toward higher use of adversatives. The slightly higher densities of additive and temporal conjunctions in Marianne's dialogue are not

sufficient to merit even the "slight tendency" label, but one expression from each of these categories, namely *or* and *when*, do show slight tendencies toward difference (see Table 1).

Let us now move from the conjunction hypotheses to the involvement hypotheses.

- **H4: Density of involvement**

This hypothesis suggests that Marianne's dialogue will be characterised by a higher density of involvement features. There was, however, no support for this hypothesis whatsoever. As can be seen in Table 2 the overall densities of these features were very close, at 211,6 per 1000 words for Elinor and 218,0 for Marianne. As noted above, the features that make up Biber's conception of involvement are key markers of face-to-face conversation in general and clearly there is no overall sense in terms of which Marianne's dialogue is more "involved" than Elinor's. Do the sisters, however, exploit the various features of involvement to different extents? This is the question that the last of the hypotheses addresses.

- **H5: Patterns of involvement**

This hypothesis suggests that the density of the 15 selected features of involvement relative to one another will differ between the two sets of dialogue. This was found to be very definitely the case, the Chi^2 test revealing a very highly significant difference ($p \leq 0,001$). Table 2 presents the frequencies and densities for each of the 15 features analysed as well as all the individual expressions which registered at least a slight tendency toward differentiation between the two dialogue corpora. To throw more light on these differences and their role in distinguishing character, the features and expressions need to be considered more closely.

This is done mainly in the sequence given in Table 2, where features are listed in descending order of their strength of association with involvement (Biber, 1988:102).

Table 1: Conjunctives

	Elinor frequency	**Marianne frequency**	**Elinor density/1000**	**Marianne density/1000**
TOTAL CONJ'S	310	226	34,8	37,0
Total dialogue	(8905)	(6104)		
ADDITIVE	102	80	11,5	13,1
or*	6	9	0,7	*1,5
and	96	71	10,8	11,6
ADVERSATIVE*	101	83	11,3	*13,6
but*	76	66	8,4	*10,8
yet*	6	10	0,7	*1,6
however*	9	2	1,0*	0,3
though	9	3	1,0	0,5
on the contrary	1	1	0,1	0,2
although	0	1	0	0,2
CAUSAL*	87	41	***9,8	6,7
for***	20	4	***2,2	0,7
if	42	26	4,7	4,3
then (causal)	9	3	1,0	0,5
because	5	6	0,6	1,0
therefore	5	2	0,6	0,3
that (causal)	5	0	0,6	0
while (causal)	1	0	0,1	0
TEMPORAL	22	22	2,5	3,6
when*	9	12	1,0	*2,1
while (temporal)	7	3	0,8	0,5
since	2	5	0,2	0,8
till	4	2	0,4	0,3

*   p≤0,25   **   p≤0,10   ***   p≤0,05

Table 2: Involvement features

	Elinor frequency	Marianne frequency	Elinor density/ 1000	Marianne density/ 1000
TOTAL INVOLVEMENT **Total dialogue**	1884 (8905)	1331 (6104)	211,6	218,0
PRIVATE VERBS*	207	121	*23,3	19,8
COGNITION VERBS****	169	76	****19,0	12,4
PERCEPTION VERBS***	38	45	4,3	***7,4
believe***	23	5	***2,6	0,8
think***	38	12	***4,3	2,0
see***	10	22	1,1	***3,6
suppose**	20	5	**2,2	0,8
mean**	14	4	**1,6	0,7
understand*	8	1	*0,9	0,2
2ND PERS PRON*****	402	165	*****45,2	27,0
your*****	97	28	*****10,9	4,6
you*****	282	130	*****31,7	21,3
"DO" AS PRO-VERB	9	7	1,0	1,1
ANALYTIC NEGATION	146	86	16,4	14,1
DEM PRO***	35	41	3,9	***6,7
this**	17	21	1,9	**3,4
GEN EMPHATICS***	46	15	***5,2	2,5
certainly**	17	4	**1,9	0,7
1ST PERS PRON*****	475	476	53,4	*****78,4
me*****	57	78	6,4	*****12,8
my*****	68	84	7,6	*****13,8
I****	276	243	31,0	****39,8
we****	26	37	2,9	****6,1

Table 2 (cont.): Involvement features

	Elinor frequency	Marianne frequency	Elinor density/ 1000	Marianne density/ 1000
PRONOUN "IT"	160	115	18,0	18,9
CAUSATIVE SUB	5	6	0,6	1,0
INDEF PRON	43	36	4,8	5,9
nothing**	17	21	1,9	**3,4
GENERAL HEDGES	9	5	1,0	0,8
AMPLIFIERS**	71	32	**8,0	5,2
very*	52	25	*5,8	4,1
WH-QUESTIONS**	42	50	4,7	***8,2
who**	4	9	0,4	**1,5
what**	14	20	1,6	**3,3
POSS MODALS	138	105	15,5	17,2
cannot**	17	22	1,9	**3,6
may*	38	16	* 4,3	2,6
NON-PHRASAL "AND"	96	71	10,8	11,6

* p≤0,25   ** p≤0,10   *** p≤0,05   **** p≤0,01   ***** p≤0,001

Private verbs are those that refer to non-observable acts and states, and include verbs of perception (e.g. *see, hear, feel*) and verbs that refer to cognition, whether intellectual states (e.g. *know, believe*) or intellectual acts (e.g. *think, assume*). The statistics show a slight tendency toward more private verbs on the part of Elinor, but if the verbs of perception and verbs of cognition are considered separately, Elinor's dialogue reveals a highly significant preference for the latter and Marianne's a significant preference for the former. With regard to individual expressions, Elinor uses significantly more of the cognition verbs *believe* and *think*, with two further such verbs, namely *suppose* and *mean*, showing a tendency toward difference and *understand* a slight tendency. Marianne, in contrast, uses *see* significantly more often than Elinor. Thus is the theme of head and heart echoed in the sisters' clashing choices of private verbs, which set Elinor's cognitions against Marianne's perceptions.

The biggest clashes of all, however, are seen in the choice of **personal pronouns**, in terms of which the sisters are neatly juxtaposed. Elinor uses *your* and *you* more often than Marianne to a very highly significant degree, with precisely the opposite applying to the preference for *me* and *my* on the part of Marianne, whose use of *I* and *we* is also very significantly higher than Elinor's. Marianne's predilection for the first-person pronouns is unusually strong, not only relative to Elinor but also to authentic modern face-to-face conversation (57,9 per 1000 words in Biber's (1988:264) corpus, as opposed to Marianne's 78,4). This finding, particularly in the singular forms (*I* is the most common word of all in her dialogue, but only the third most common in Elinor's), corroborates and helps to explicate critics' views, such as that Marianne is "sick with the intensity of her own passions" and tends to "make language a monologue and behaviour a monodrama" (Brown, 1979:58 and 60). The finding on her very sparing use of *you* (only the sixth most common word, compared to the second most common – after *to* – with Elinor) helps similarly to explain, for example, why her very "sensibility" has been seen as a source of her "thoughtlessness" and her "unwillingness to sense the feelings of others" (ApRoberts, 1986:50). Elinor's preference for second over first person pronouns, on the other hand, reflects her much more "extended social consciousness" (ApRoberts, 1986:50).

The two dialogues reveal very similar frequencies of Biber's involvement features **DO as pro-verb** and **analytic negation** (the use of *not* rather than synthetic forms of negation such as neither and *none*), but the next feature on Biber's list, **demonstrative pronouns**, reveals a significant difference in favour of Marianne. Of the individual demonstrative pronouns, however, only *this* shows a tendency toward differentiation. It is often used to refer not to a specific nominal elsewhere in the discourse but rather, more vaguely, to complexes of ideas and events in the discourse context, frequently occurring in emotionally charged language. Typical, thus, are utterances of Marianne's such as *Elinor, ... is **this** fair? is **this** just? are my ideas so scanty?* (Austen, 1959:71) and *Good God! Willoughby, what is the meaning of **this**?* (Austen, 1959:192), cried out at the climactic moment when Marianne confronts the man who has, since early in the novel, aroused in her a confusion of the most passionate emotions. Here the *this* is powerfully evocative of all that she has suffered since his sudden departure of many months earlier.

General emphatics (e.g. *definitely, certainly*, *indeed*) emphasise speaker commitment to the certainty of propositions they express and here Elinor reveals a significantly higher use overall. One of this class of expressions, *certainly*, shows a tendency toward significantly higher use.

It could be argued that Elinor's higher rating here links with her "coolness of judgement" and self-assurance.

Very similar densities of use of the **pronoun *it*** are found in both dialogues. This similarity applies also to Biber's category of **causative subordination**, which he limits to the use of *because* (see Table 1 for statistics on this expression); to **general hedges** (markers of probability or uncertainty, such as *almost, more or less*); and to **indefinite pronouns** (e.g. *anyone, no-one, something*), though one expression in this category, *nothing*, shows a tendency toward higher use on the part of Marianne. This can be partly explained by the emotional power of the word as used in some of her utterances, such as *Mine is a misery which **nothing** can do away* and later – more positively – *... and I have **nothing** to regret – **nothing** but my own folly* (Austen, 1959:201 and 357).

Elinor shows a tendency towards higher use of **amplifiers** (e.g. *entirely, highly, very*), mostly because of a slight tendency to favour *very*, the most frequent amplifier in both dialogues. Amplifiers mark a strong positive conviction toward the proposition expressed, and relate closely to general emphatics, so it is not surprising that we see here a similar pattern of differentiation between the sisters on this feature.

A significantly higher use of **WH-Questions** characterises Marianne's dialogue, the main contributing factors here being the tendencies toward a preference for *who* and *what*. A predilection for such questions reflects Marianne's being "eager in every thing", including being prone to string questions together when in distress, as in *And yet this woman – **who** knows what her art may have been ... **Who** is she? **Who** can she be?* (Austen, 1959:206).

Although the density of use of **possibility modals** (in Biber's [1988:241] terms, all uses of *can, could, may* and *might*) in the two dialogues was similar, there was a degree of variation with regard to individual expressions. Marianne reveals a tendency toward a significant preference for *cannot*, mostly – unlike Elinor – together with "*I*" as subject, where this negative, as in Marianne's use of *nothing* (see above), tends to signal emotional involvement, for example in *I **cannot** stay here long, I **cannot** stay to endure the questions and remarks of all these people* (Austen, 1959:206). Elinor, on the other hand, reveals a slight tendency toward a preference for *may*, a modal regularly used as a hedge, including in careful argumentation, as in "Willoughby *may* undoubtedly have very sufficient reasons for his conduct, and I will hope that he has. But it would have been more like Willoughby to acknowledge them at once" (Austen, 1959:102).

The final involvement feature is **coordination of non-phrasal units by and**. This feature was dealt with as a subcategory of additive conjunction earlier and revealed no tendency toward differentiation between the two dialogues.

6. Conclusions

The fine-grained analysis of dialogue undertaken here shows that the sisters' linguistic behaviour differs in definite but subtle ways, and that some of these differences can be linked to the opposition between sense and sensibility. There was no support for Hypotheses 1 and 4, indicating no preference by Elinor for adversative conjunctions and no sign of Marianne's conversation being generally more "involved" than Elinor's. Elinor does, however, use significantly more causal conjunctives (Hypothesis 2) and the sisters differ significantly in their patterns of use of both conjunctives and of involvement features (Hypotheses 3 and 5).

Moreover, counts of features and of individual expressions (as given in Tables 1 and 2) reveal that the dialogues differ at least significantly (sometimes highly significantly or very highly significantly) in terms of two conjunction parameters and 16 involvement parameters. It should be mentioned here that in order to test further the implication that these findings show that the sisters' language behaviour is both distinctive and individually consistent, a supplementary count was made comparing the first half of each character's dialogue with the second half. The only statistically significant (p≤0,05) differences in these counts were that Elinor uses the pronoun *it* more in the second half of her dialogue and that Marianne uses more first person pronouns, and specifically *me, my* and *I*, in the second half of hers. While reasons for the change in Elinor's use of *it* are not readily apparent, Marianne's higher use of the first person singular pronouns in the second half of her dialogue can be attributed in considerable degree to the fact that the beginning of the second half coincides closely with her confrontation with Willoughby, after which she becomes still more self-obsessed, and until her (ultimately liberating) illness tries to avoid the company of all except her sister-cum-confidante. Other corpus-based studies could trace how dialogue features might shadow perceived changes in characters, or their situations, but here the supplementary split-half comparisons serve to reveal remarkable consistency in the linguistic behaviour of both Elinor and Marianne, with the number of significant differences between them far outweighing inconsistencies internal to their dialogue corpora.

Many of the specific differences between the sisters can be linked to the more explicit authorial and non-authorial character cues (see Section 2 above) provided in the novel, which readers and critics would be more

conscious of. The scrutiny undertaken here of implicit cues provided in the dialogues, and of their distinctive statistical patterning, thus reinforces literary perceptions and goes some way towards explicating them. It is therefore hoped that this study gives credence to Burrows' (1987: 106) claim that "there is obviously room for a closer working relationship between the literary critic, the statistician, and the student of linguistics than is common nowadays".

Bibliography

ApRoberts, R. 1986. Sense and Sensibility, or Growing up Dichotomous. In: Bloom, H. (ed.) *Jane Austen*. New York : Chelsea House. p. 43-55.

Austen, J. 1959. *Sense and Sensibility*. New York : Dell.

Biber, D. 1988*. Variation across Speech and Writing*. Cambridge : Cambridge University Press.

Brown, J.P. 1979. *Jane Austen's Novels*. Cambridge, Mass. : Harvard University Press.

Burrows, J.F. 1987. *Computation into Criticism: a Study of Jane Austen's Novels and an Experiment in Method*. Oxford : Clarendon.

CEOED 1971. *The Compact Edition of the Oxford English Dictionary*. Oxford : Clarendon.

Chatman, S. 1972. On the formalist-structuralist theory of character. *Journal of Literary Semantics*, 1:57-79.

Culpeper, J. 2001. *Language and Characterisation: People in Plays and Other Texts*. Harlow : Pearson Education.

Fahnestock, J. 1983. Semantic and lexical coherence. *College Composition and Communication*, 34(4):400-416.

Ferguson, G.A. 1981. *Statistical Analysis in Psychology and Education*. Singapore : McGraw-Hill.

Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London : Longman.

Hatch, E. & Farhady, H. 1982. *Research Design and Statistics for Applied Linguists*. Rowley : Newbury House.

Kruger, A. 2000*. Lexical Cohesion and Register Variation in Translation: "The Merchant of Venice" in Afrikaans*. Pretoria : University of South Africa. (D.Litt. et Phil. thesis.)

Morton, A.Q. 1978. *Literary Detection*. Bath : Bowker.

Short, M.H. 1996. *Exploring the Language of Poems, Plays and Prose*. Harlow : Addison Wesley Longman.

Short, M., Semino, E. & Culpeper, J. 1996. Using a corpus for stylistics research: speech and thought presentation. In: Thomas, J. & Short, M. (eds.) *Using Corpora for Language Research*. Harlow : Longman. p. 110-131.

Stokes, M. 1991. *The Language of Jane Austen*. London : MacMillan Education.

Tannen, D. 1985. Relative focus on involvement in oral and written discourse. In: Olson, D.R., Torrance, N. & Hildyard, A. (eds.) *Literacy, Language and Learning*. Cambridge : Cambridge University Press. p. 124-147.

Van Peer, W. (ed.) 1989. *The Taming of the Text: Explorations in Language, Literature and Culture*. London : Routledge.

Watson, G. 1999. Something will come of nothing: an empirical model of negation for prose literature. *Journal of Literary Studies*, 15(3/4):377-406.

Williams, M. 1986. *Jane Austen: Six Novels and their Methods*. London : Macmillan.

Zyngier, S. 1999. 'Smudges on the canvas'? A corpus stylistics approach to Macbeth. In: Biermann, I. & Combrink, A. *Poetics, Linguistics and History: Discourses of War and Conflict.* Potchefstroom : Potchefstroom University. p. 529-545.

Key concepts:

Austen, Jane: characterisation; dialogue
corpus linguistics
linguistic stylistics
text linguistics

Kernbegrippe:

Austen, Jane: karakterisering; dialoog
korpuslinguistiek
linguistiese stilistiek
tekslinguistiek