
Can emotional valence in stories be determined from words?¹

Yves Bestgen

*Fonds national de la recherche scientifique
Université catholique de Louvain
Place du Cardinal Mercier, 10 B1347 Louvain-la-Neuve Belgique
yves.bestgen@psp.ucl.ac.be*

ABSTRACT. In spite of the growing interest witnessed in the study of the relationship between emotion and language, the determination of the emotional valence of sentences, paragraphs or texts has, yet, attracted little attention. To bridge this gap, a technique based on the emotional aspect of words is presented. In this preliminary study, we have compared the affective tones of the sentences of four texts as perceived by readers, to the values generated by the words that compose the texts. The results support the psychological reality of the affective tones of linguistic units larger than a word, and the possibility of their evaluation through the lexical information. Such information should be useful for studying the role of emotional interest on text processing and for the analysis of the natural stories produced by people in reaction to stressful events.

This paper deals with the problem of determining the emotional valence of verbal stimuli. It has long been resolved in the case of the words by asking judges to evaluate them on scales like "pleasant — unpleasant" (Osgood, Suci & Tannenbaum, 1957; Heise, 1965). Such normative word lists were fruitfully used in numerous studies on the role of emotional stimuli on cognitive processing. However, the emotional content of longer linguistic units such as sentences or paragraphs was continuously neglected in spite of its importance for the study of the verbal expression of emotion. This neglect is primary due to a methodological lack that can be summarized by the following question : How can one determine the emotional content of units larger than a word? A possible technique for answering this question was proposed by Anderson and McMaster (1982) and a few others (Hogenraad & Bestgen, 1989; Martindale, 1987; Whissell & Dewson, 1986). After presenting two justifications for a study of the emotional content of text, I will describe this technique and give some empirical support for its suitability.

¹ This work was supported by grant 1.5.140.91F (FNRS) to Yves Bestgen (Senior Research Assistant, Belgian National Fund for Scientific Research).

Recent studies about people's reactions to stressful events highlight the necessity of a tool that allows the determination of affective tone in discourse. In a study on the social sharing of emotion, Rimé, Mesquita, Philippot and Boca (1991) showed that people who had faced an important emotional event, initiated communicative processes to share their experience with others. In its strongest form, social sharing corresponds to what Harvey, Orbuch and Weber (1990) called "account in response to severe stress," that are story-like explanations for high emotional events. Such accounts can be observed during the successive steps of a stress response sequence that conducts a person from a traumatic event to the completion state (acceptance). Because these accounts are primary verbal, further developments in this research area depend on the possibility to quantify their emotional content. Such a tool would allow the analysis of the emotional content of the successive accounts made by a person, but also the comparison of the impact of various stressor events such as disease or criminal victimization.

The need of a tool for determining the affective content of linguistic units can also be justified by an analysis of the interest produced by a text. A growing number of researchers have been disappointed by the limitations of story grammars and other cognitive models of discourse comprehension, and now consider that affect plays a sizable role in reading, comprehension and recall of texts. The influence of affect, long masked by the extreme simplicity of the stories analyzed by story grammars (Miall, 1989), is felt in its participation in determining the level of interest in a text (Hidi & Baird, 1986). Interest has traditionally been characterized by two components: the cognitive component and the emotional one. The first is discussed by authors like Kintsch (1980) who suggested three factors to explain it: reader's background knowledge, unpredictability of the following parts of a text, and resolution of the uncertainty at the end of the text. Experimental supports for these factors were presented by Iran-Nejad (1987) and Garner and Gillingham (1991).

The emotional component is produced by the very content of a text. It is quoted in numerous theoretical propositions: themes of absolute interest (Schank, 1979), events that are emotional in context as well as out of it (Kintsch, 1980), point or human dramatic situations (Wilensky, 1983), significant events that could have important consequences for one character (Brewer & Lichtenstein, 1981). The most serious objection faced by this component is the lacks of an objective and empirical definition (see, for instance, some of the peers' commentaries to Wilensky's provocative 1983 paper in which he developed his "point" conception of a story — Allen, 1983; Mandler, 1983; Marshall, 1983; Walz & Dorfman, 1983). To adopt an integrated approach of interest and storiness, as advocated by Stein (1982, 1983), structural model like story grammar and cognitive interest should be supported by a method of text analysis that would allow the quantification of the emotional content in stories.

The methodological objective of this study is to present and to validate two procedures to estimate the emotional intensity of verbal stimuli. Asking judges to rate such stimuli should be a viable solution as long as the level of agreement

between them is high. Only a few studies gave some tentative data about this question and their results seem unclear. Sadoski and his colleagues asked 15 subjects to evaluate their emotional reaction to paragraphs from three stories (Sadoski, Goetz & Kangiser, 1988) and from three magazine articles (Sadoski & Quast, 1990). The reliability of the average scores obtained from responses from all judges are all above 0.92. Miall (1989) asked 12 judges to evaluate the affective intensity of the phrases in the introduction to one of V. Woolf's short stories, a text that contains a number of highly difficult point of interpretation. Although highly significant, the reliability coefficient reached a value of only 0.56. A positive answer to this interrater agreement is needed before one can use judges' advice to determine the affective content of linguistic units larger than a word.

But, even if this requirement is fulfilled, the use of judges is only applicable when compiling experimental material composed of a limited number of sentences or paragraphs. On a large scale, the intrinsic combinatory nature of language renders this solution unpractical. Contrasting what happens when such rating procedure is used with words or with sentences could be useful. Generating a list of the most frequently used words of a language, or indeed, of all of its words, then obtaining their emotionality scores is possible. Any similar attempt for sentences is futile. Each new sentence, in fact probably each sentence, has to be rated by a series of judges. It would, therefore, be advantageous to be able to determine the emotional content of a text based on the words that compose it. This procedure was originally proposed by Heise (1965) who used a list of 1000 words rated according to three semantic differential dimensions (evaluation, activation, and power; Osgood, Suci & Tannebaum, 1957) in order to quantify the verbal productions of subjects. Anderson and McMaster (1982, 1986) expanded this proposition to model the affective tone within texts. In a series of exploratory analyses, they applied this technique to various chapters of novels and to children's stories, showing that these texts present an "emotional crisis" which can be linked to the idea of dramatic progression: "An ebb and flow of emotional tension, traditionally expected to build toward a gratifying release near end of well-crafted stories." (Anderson & McMaster, 1982, p.1).

The proposed technique is relatively straightforward. First, the text is divided into segments (length is unimportant), and a list of all the different words and their frequency within each segment is drawn up. Then, this list is compared to a dictionary containing the words for which we know their degree of evaluation, activation, or power (for example: Heise's proposed list); each time a word is found to be on both lists, its score is recorded. Finally, the average score for each segment is calculated.

Until now, this technique has received little attention due probably to its simplistic approach. It examines the words completely out of context, without taking into account, for example, negations or quantifiers, which highly modify the affective values of combined words (Cliff, 1959; Howe, 1963). Moreover, it uses only a sampling of words that compose the text, rather than all of them. These

elements underscore the need for repeated validation. Besides from the illustrative application of this technique to different texts, the only conclusive evidence comes from Anderson and McMaster (1982). They compared texts written by students about the best and worst possible future scenarios they could imagine, and found that the evaluation and activation scores were significantly higher for the texts describing the best possible scenarios. However, the validity of this technique for describing the affective tone within a text was never demonstrated. One of the goals of the present study is to collect further information on this validity. But we want to go further than this instrumental point of view. The study of the emotional intensity of verbal stimuli can lead to the study of the relationship between emotion and language by showing that readers are sensitive to, and are able to assess, this parameter.

Subjects were asked to read a text, then evaluate the emotional valence (pleasant-unpleasant) of each sentence. The analysis of the inter-judge agreement allowed us to verify that the sentences displayed an affective tone perceived by the majority of the subjects. We then calculated the emotional valence of each sentence based on its words. The magnitude of the correlation between these two indices allowed the validation of the technique.

However, this method of testing our hypothesis didn't allow us to control an important parameter: the role of text sequence. The subjects read the sentences within their context, whereas the analysis based on the words could take into account neither the macro-context of discourse, nor the micro-context of the sentence. It is possible that the text sequence should guide, in part, their evaluations. Therefore, we obtained also the opinions of "reader-judges" on the emotional nature of these sentences out of context. This information gave to the study a context gradient: word, sentence, text.

This study is essentially exploratory, and its generalization is limited for at least two reasons. First, we used a small number of texts, all short stories. This is primarily because affective valence seems to be an important dimension in short stories (Anderson & McMaster, 1982; Wainer & Berg, 1972), and because preceding studies on this subject also used them. Moreover, it is noteworthy that people's accounts of traumatizing experiences share the characteristics of short stories (Harvey et al., 1990; Stein, 1982). So, a valid technique for short stories would probably be useful for the analysis of accounts.

Furthermore, we only studied the main dimension of emotions: evaluation. We decided against the use of a general scale of "non emotional — very emotional" because, this did not allow us to differentiate between valences of, for example, "great joy" and "great sorrow". But this does not mean that other dimensions as these studied by Anderson and McMaster (1982, 1986) or Hogenraad and Bestgen (1989) are without any interest.

Method

Material

Our goal, in selecting the material to be used, was to cover several literary styles: the classic tales, modern literature stories, and short stories. Four texts were chosen: "The Little Match Girl" by Andersen (AND; 1987), "He Belonged to Me, Said the Sea" by Cesbron (CES; 1959), "The Seven Ravens" by the brothers Grimm (GRI; 1976), "The Stroll" by Maupassant (MAU; 1974). "The Seven Ravens" is a traditional tale, whereas "The Little Match Girl" doesn't retain this classic style. Both are considered children's literature. The remaining two texts are for older readers: "The Stroll" is a short story by one of the greatest nineteenth century French authors, and "He Belonged to Me, Said the Sea" is a story, or fantasy, by a modern author. The narrative structures of these texts can be seen in the synopsis (Appendix 1).

The four texts were divided into sentences (segments), generally according to the periods. In some cases, however, long sentences were sub-divided into several shorter ones (according to semi-colons), and particularly short ones were grouped into a longer one. Table 1 indicates the length of each text, in number of words and segments, and the average length (and standard deviation) of a segment, in number of words.

	AND	CES	GRI	MAU
Number of segments	38	34	34	92
Number of words	971	700	936	1984
Average number of words per segment	25.55	20.59	27.53	21.57
Standard deviation	15.64	13.19	10.41	10.26

Table 1: Descriptive statistics for the 4 texts.

Tasks

Text Task

The texts were divided into two groups that contained approximately the same number of sentences to judge. The first group consisted of the three shortest texts, and the second consisted of "The Stroll" by de Maupassant.

Each story was divided into sentences and these were presented on several large pages. The text was on the left side of the page, and on the right side, for each sentence was a 7-point rating scale. The top line of each page consisted of the title of the story, the author's name, and on the far right of the page were the two limits of the scale: unpleasant and pleasant. These limits were repeated at the bottom of each page. Each subject was told that we were interested in the affective content of the texts, and that we wanted to know if different judges would perceive the same affective progression. For each segment, we asked the subjects whether the situation described evoked an unpleasant idea or a pleasant one. We insisted that their judgments indicate the progression perceived throughout the text. Finally, the 7-point rating scale was given verbal equivalents: very (unpleasant — pleasant), medium, a little, and neutral.

Sentence Task

The goal in this task was to obtain an evaluation of the pleasantness of the sentences without the subject having access to the narrative context. The text sequences were broken up in the following manner: all sentences, from all four stories were divided into three sets, and these sets were prepared for three groups of subjects. Each set contained one third of all the sentences of each text. More precisely, each story was divided into three packs of sentences by choosing systematically one sentence in three. Then, the three sets were assembled by randomly choosing one pack of sentences from each story. Within each of these sets, the sentences were placed in a random order, with the only constraint being that no two sentences from the same story could directly follow one another.

In this task, each sentence was presented on a small sheet, and below each sentence was the 7-point rating scale with the same limits as in the preceding task. The sentences that the subject had to judge were given in two numbered booklets. The subject was asked to read each sentence and evaluate it, before going to the next one.

A pretest showed that the main difficulty in this task was due to the lack of anchoring of the scale. Indeed, the subject may not read all the sentences before starting the evaluation, and therefore may have some difficulties in situating the first few sentences regarding the extremes of the scale. To resolve this problem, we added ten sentences at the beginning of each set. These sentences were taken from other stories, and were chosen so that every point on the rating scale would be represented, from the "very unpleasant" to the "very pleasant". The subject was not made aware of this, and these extra sentences were disregarded in the analysis.

Lexical Task

A list of all the different words from the four texts, was drawn up, although we did not take into account articles, pronouns, prepositions, proper names, and the verbs "to be" (*être*) and "to have" (*avoir*); all words that, in this case, weren't worth evaluating (Heise, 1965). The list contained 840 different words, which corresponds

to an average number of words taken into account per text, between 40% (for CES) and 43% (for AND). The 840 words were distributed randomly into three questionnaires. The words were on the left side of the page, and on the right side of the page, was the 7-point rating scale of "unpleasant — pleasant". As in the preceding task, 10 words were added to the beginning of each list to anchor the scale. The instructions given to the subjects were similar to the ones in the two previous tasks.

Subjects

Principal Sample Group

One hundred twenty subjects took part in the experiment. Each type of questionnaire (two for the first task, three for the two others) was answered by 15 subjects. All subjects were male, and French was their mother tongue. They were students at the Royal School of the National Police Force (Ecole Royale de Gendarmerie; Brussels, Belgium), and were all between the ages of 18 and 22 years. The questionnaires were answered by the students in one sitting.

Control Group

Because of the particular type of principal sample group used, we had to verify the generalization of their interpretations with a less specific population. Four subjects, all female graduate students, did the text task for the three shorter texts and five students, four females and one male, did the sentence task.

Concerning the third task, new evaluations were not necessary because, amongst the 840 words retrieved from the texts, 416 of them were already in Leleu's dictionary (Hogenraad & Bestgen, 1989; Leleu, 1987), and 161 were on the list drawn up by Messina, Morais, and Cantraine (1989). Leleu's list is composed of 2999 words that have been evaluated by an average of 39 judges, all students in high school or in technical schools. As for Messina's et al. list (1989), it is composed of 904 words that have been evaluated by 209 judges between the ages of 17 and 90 years. We will, therefore, be able to compare our results directly to these two lists.

Results

The principal goal of this study is to compare three different ways of obtaining the affective valence of sentences that compose a text. Before comparing these scores, two analyses must be carried out. First, the inter-judge agreement must be calculated. If, for one or more of the tasks, it is found to be too small, we would not be able to compare the average tendency of each group with the others. Next, we have to compare the responses of the principal sample group to those of the control group.

Inter-judge agreement

In order to estimate the inter-judge agreement, the correlation was calculated between all possible pairs of judges within each group. These correlation matrices were summarized by calculating the mean correlation between all the subjects and by executing a principal components analysis. The mean correlations are 0.55 for the textual task, 0.49 for the sentential one, and 0.50 for the lexical one. For each task, the first factor accounts for half of the variance or more, whereas the second factor accounts for only a small percentage of it (8% or less). Finally, the "Theta" coefficients, which assess the reliability of the factor scores (Armor, 1974; Rosenthal, 1982), are all above 0.93, showing that the factorial scores can be trusted.

It is worth noting that the inter-judge agreement scores obtained in the first two tasks are equivalent to those obtained in the third task, the Lexical one. Moreover, the reliability of the three tasks is equivalent to the values observed with a lexical task by other researchers (Brown and Ure, 1969; Messina et al., 1989; Silverstein and Dienstbier, 1968).

Comparison of the two Sample Groups

In the first two tasks, the inter-judge agreement scores were approximately equivalent for both the Principal Sample Group and the Control Group (mean r of 0.55). In order to compare their evaluations, we correlated the factorial scores derived from each of the two sampling groups. This value is 0.87 for the Text Task and 0.88 for the Sentence Task, both significant at $p < 0.001$.

In the Lexical Task, we calculated the correlation between the factorial scores obtained, and the scores for the 413 words found in common, with Leleu's dictionary (1987); the correlation is 0.93 ($p < 0.001$). The same method was used for the 161 words in common with Messina's et al. list (1989); the correlation is 0.87 ($p < 0.001$).

Estimation of the affective valence of the sentences based on the words

For the text and the sentence task, estimation of the affective valence of the sentence was straightforward; we used the factorial scores as derived during the analysis of the inter-judge agreement.

In order to determine the affective score of the sentences solely based on the words, we used the PROTAN software (Hogenraad, Daubies, & Bestgen, 1988) that, among other things, compares the words of a text to those in a dictionary. When it finds a match, it attaches to that word its corresponding affective score, the factorial score of the word deduced from the lexical task. The score for a sentence is none other than the average of the scores of all the words in that sentence.

A coder analyzed each text to take into consideration the ambiguity of certain words or the negations. First, he eliminated the scores of the words when their meaning was not the same in the text and in the dictionary. Next, the affective scores for the negated words were reversed. To guarantee the reliability of this coding, one of the texts was re-analyzed in the same manner three months later by the same judge, as well as by another independent judge. The correlations between the affective values resulting from the three analyses are equal or superior to 0.98.

Comparison of the affective tones in the three tasks

The three affective tones of the sentences were compared using correlation coefficients. Table 2 shows that the correlations between the textual and sentential tones are similar for all four texts, and have very high levels of significance. What about the correlations between the lexical tones and the first two? If these are inferior to the correlations between the textual and sentential tones, then they are superior to the mean inter-judge correlations, and have also very high levels of significance.

	Textual Sentential	Textual Lexical	Sentential Lexical
AND	0.77	0.74	0.77
CES	0.77	0.72	0.65
GRI	0.80	0.61	0.55
MAU	0.84	0.60	0.69

Table 2: Correlations between the textual, sentential, and lexical profiles for the 4 texts.

We must now consider the context gradient. Are the lexical values more closely linked, as suspected, to the sentential than to the textual values? This hypothesis is indeed verified for two texts (AND and MAU), but invalidated for the other two. These pairs of correlations are significantly different for only one text: "The Stroll" by de Maupassant ($Z = 2.12$; $p < 0.05$; comparison of correlated correlation coefficients: Meng, Rosenthal, & Rubin, 1992). This sole significant difference is insufficient to confirm our hypothesis. Moreover, we can show, with a multiple regression analysis (with the textual values as the dependent variable and the sentential and lexical values as the independent variables), that the lexical scores are linked to the textual scores, and this independently from the sentential scores. Indeed, three times out of four, the beta-weight for the lexical scores is significant

(AND : $\beta=0.36$, $t(35)=2.29$, $p<0.03$ — CES : $\beta=0.38$, $t(31)=2.83$, $p<0.01$ — GRI : $\beta=0.25$, $t(31)=2.11$, $p<0.05$). "The Stroll" is the only text for which such a link was not significant ($\beta=0.03$, $t(89)=0.35$, $p>0.05$). The existence of a direct relationship between lexical and textual affective values is difficult to explain, because the subjects who evaluated the sentences saw the words, just as did the subjects that evaluated the texts. It is as if the sentential context sometimes contradicts the lexical information, despite the rehabilitation of this information by the text.

Overall, the scores derived from the dictionary of emotional scores are very significantly linked to the valences perceived by the readers. This is an argument in favor of the validity of the proposed procedure; that is, the use of a dictionary that is exhaustive as well as adapted to the texts to be analyzed. But what happens when these two elements are ignored, such as in the analyses in which only a random sample of the words of a text is used without any attention to negation (Anderson & McMaster¹, 1982, 1986, 1989; Hogenraad & Bestgen, 1989; Whissell & Dewson, 1986)? To answer this question, we determined the affective profiles of the texts using a partial and un-adapted dictionary (Leleu's dictionary, 1987, 2999 words), an exhaustive but un-adapted dictionary (the one we put together, but without the extra coding), and finally a partial dictionary that was adapted to the texts (Leleu's adapted dictionary). Table 3 presents the correlations between the different profiles and the readers' evaluations.

Table 3: Correlations between the lexical affective profile and the textual and sentential affective profiles for the 4 texts.

	Partial Un-adapted	Exhaustive Un-adapted	Partial Adapted
AND-Textual	0.22	0.38*	0.60***
AND-Sentencial	0.40*	0.61***	0.62***
CES-Textual	0.41*	0.63***	0.46**
CES-Sentencial	0.38*	0.59***	0.42*
GRI-Textual	0.29	0.35*	0.48**
GRI-Sentencial	0.24	0.34*	0.39*
MAU-Textual	0.27**	0.38***	0.53***
MAU-Sentencial	0.37***	0.54***	0.60***

Note: Levels of significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$. The interpretation of the labels is indicated in the text.

It is apparent that the classic procedure is not very effective. The use of an exhaustive dictionary greatly improves the correlations. Finally, an analysis based on a partial dictionary, but adapted to the texts, allows a significant prediction of the affective tones perceived by the readers. Nevertheless, these correlations are, six times out of eight, significantly lower than those observed between affective profiles determined by the subjects and the exhaustive and adapted dictionary (Table 2). The only non-significant comparisons are with the text task in "The Seven Ravens" ($Z=1.59$; $p<0.06$) and in "The Stroll" ($Z=1.10$; $p<0.15$) — unidirectional test. Of course, these results are exclusively valid for the dictionary used (2999 words evaluated by an average of 39 judges).

Discussion

The goal of this study was twofold. First, we wanted to demonstrate that linguistic units larger than a word do indeed incorporate affective tones detectable by the readers. Second, we wanted to study the possibility of predicting these tones using the affective content of the words in the text. By answering these questions, we hoped to show that Anderson and McMaster's technique can be used extensively and not simply exploratory as it was the case in the studies published until now by these authors (1982, 1986) and others (Hogenraad & Bestgen, 1989; Martindale, 1987).

Although the inter-judge agreement scores are not perfect, we have been able to show, for each text, a common tendency perceived by the majority of subjects, which would explain a large number of individual scores. This brings us back to Sadoski's et al. observations (1988, 1990) concerning emotionality. The present experiment also demonstrates that the affective tone of sentences is relatively independent of the textuality. Moreover, our study, as well as Sadoski et al. and Miall's study (1989), all have the same bias: in the Text Task, the subjects first read the entire text. It is therefore possible that knowledge of the outcome of the story interferes with the affective score given to a particular sentence and the preceding ones, and reduces the similarity between the Text task and the Sentence Task. It would be interesting to compare the two conditions that we used, to a third one in which the subject would read each sentence of the text, one by one and in order, and would evaluate each one before going on to the next one (Miall, 1990). This situation would probably yield intermediate scores between the sentential and the textual scores as defined in the present study.

The results of determining the affective structure of a text from the words that it is composed of, are encouraging. This study highlights the necessity to use an adapted and exhaustive dictionary to reach such efficiency in opposition to what was done in older studies (Anderson & McMaster, 1982, 1986, 1989; Hogenraad & Bestgen, 1989; Whissell & Dewson, 1986). The necessity of continually perfecting the proposed technique is emphasized in that way. Other variables than negations must also be considered. For example, Mayer (1940) in a study of the affective

As one can see, some sentences are highly pleasant (near +3); others are highly unpleasant (near -3). The first cycle that leads the little girl from hopelessness to joy and then brings her back to hopelessness is obvious on the curve (sentences 14 to 22).

Facing this kind of data, one can ask whether neutral sentences would be less recalled by the reader than highly pleasant or highly unpleasant ones. Certainly, the structural importance of sentences should be taken into account. Perhaps, the most important study would be to see whether the emotional curve could be linked to "interestingness" in stories but also in expository text. This relation could not be perfect at least for two reasons. First, as said in the introduction, the analysis of affective tones can only take into account one component of "interestingness": the emotional and not the cognitive component. Furthermore, we analyze only one dimension of the emotional content, the one that seems to be the most important (Frijda, 1987) but this doesn't mean that it's the only interesting one. Anderson and McMaster (1986) proposed to use the three semantic differential dimensions to distinguish eight emotional states based on the work of Mehrabian (1980). For instance, a textual segment that would be high in evaluation, in activity and in potency, could be called "exuberant". However, before one can use this combined score, specific validations are needed.

In conclusion, the lexical approach advocated in this paper should be very useful with large corpus as short stories from various countries (Martindale, 1987) or natural discourse like accounts (Harvey et al., 1990). The high correlation between the dictionaries of Leleu (1987) or of Messina et al. (1989) and ours shows that such a list of words can be completed according to the need, as it was originally proposed by Heise (1965).

References

- Allen, J.F. (1983). Do story grammars and story points differ? *The Behavioral and Brain Science*, 4, 592-593.
- Andersen, H.C. (1987). *Contes choisis*. Paris: Gallimard.
- Anderson, C.W., & McMaster, G.E. (1982). Computer assisted modeling of affective tone in written documents. *Computers and the Humanities*, 16, 1-9.
- Anderson, C.W., & McMaster, G.E. (1986). Modeling emotional tone in stories using tension levels and categorical states. *Computers and the Humanities*, 20, 3-9.
- Anderson, C.W., & McMaster, G.E. (1989). Quantification of rewriting by the brothers Grimm: A comparison of successive versions of three tales. *Computers and the Humanities*, 23, 341-346.
- Armor, D.J. (1974) Theta reliability and factor scaling. In H.L. Costner (Ed.), *Sociological methodology, 1973-1974* (pp.17-50). San Francisco: Jossey-Bass.
- Brewer, W.F., & Lichtenstein, E.H. (1981). Event schemas, story schemas, and story

14 Final version in *Cognition and Emotion*, 7, 21-36

- grammars. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX*, (pp.363-379). Hillsdale: Lawrence Erlbaum Associates.
- Brown, W.P., & Ure, D.M.J. (1969). Fived rated characteristics of 650 word association stimuli. *British Journal of Psychology*, 60, 233-249.
- Cesbron, G. (1959). *Tout dort et je veille*. Paris: Robert Laffont.
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66, 27-44.
- Frijda, N.H. (1987). Comment on Oatley and Johnson-Laird's "Towards a cognitive theory of emotion". *Cognition and Emotion*, 1, 51-58.
- Garner, R., Gillingham, M.G. (1991). Topic knowledge, cognitive interest, and text recall: A microanalysis. *Journal of Experimental Education*, 59, 310-319.
- Grimm (1976). *Contes*. Paris : Gallimard.
- Harvey, J.H., Orbuch, T.L., & Weber, A.L. (1990). A social psychological model of account-making in response to severe stress. *Journal of Language and Social psychology*, 9, 191-207.
- Heise, D.R. (1965). Semantic differential profiles for 1000 most frequent english words. *Psychological Monographs*, 79, 1-31.
- Hidi, S., & Baird, W. (1986). Interestingness - A neglected variable in discourse processing. *Cognitive Science*, 10, 179-194.
- Hogenraad, R., & Bestgen, Y. (1989) On the thread of discourse: Homogeneity, trends, and rhythms in texts. *Empirical Studies of the Arts*, 7, 1-22.
- Hogenraad, R., Daubies, C., & Bestgen, Y. (1988). Une théorie et une méthode générale d'analyse du contenu assistée par ordinateur : le système PROTAN (PROTOCOL ANalyser). Faculté de Psychologie et des Sciences de l'Education. Université Catholique de Louvain, Louvain-la-Neuve. Document non publié.
- Howe, E.S. (1963). Probabilistic adverbial qualifications of adjectives. *Journal of Verbal Learning and Verbal Behavior*, 1, 225-242.
- Iran-Nejad, A. (1987). Cognitive and affective causes of interest and liking. *Journal of Educational Psychology*, 79,120-130.
- Kintsch, W. (1980). Learning from text, levels of comprehension, or: why anyone would read a story anyway? *Poetics*, 9, 87-98.
- Leleu, S. (1987). Un atlas sémantique de concepts d'émotion : normes et validations. Faculté de Psychologie et des Sciences de l'Education. Université Catholique de Louvain, Louvain-la-Neuve. Mémoire de licence non publié.
- Mandler, J.M. (1983). What a story is. *The Behavioral and Brain Science*, 4, 603-604.
- Marshall, J.C. (1983). Wilensky's recipe for soap-opera scripts, or Marcel Proust is a yenta. *The Behavioral and Brain Science*, 4, 604-605.
- Martindale, C. (1987). Narrative pattern analysis: A quantitative method for inferring the symbolic meaning of narratives. In L. Halasz (Ed.), *Literary discourse: Aspects of cognitive and social psychological approaches* (pp.167-181). Berlin: de Gruyter.

- Maupassant, G. de (1974). *Contes et nouvelles*. Paris: Gallimard.
- Mayer, G. (1940). *La qualification affective dans les romans d'Honoré de Balzac*. Paris: Nizet.
- Mehrabian, A. (1980). *Basic dimensions for a general psychological theory*. Cambridge: Oelgeschlager, Gunn and Hain.
- Meng, X.-L., Rosenthal, R., & Rubin, D.B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175.
- Messina, D., Morais, J., & Cantraine, F. (1989). Valeur affective de 904 mots de la langue française. *C.P.C.: European Bulletin of Cognitive Psychology*, *9*, 165-187.
- Miall, D.S., (1989). Beyond the schema given: Affective comprehension of literary narratives. *Cognition and Emotion*, *3*, 55-78.
- Miall, D.S., (1990). Readers' responses to narrative: Evaluating, relating, anticipating. *Poetics*, *19*, 323-339.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Rimé, B., Mesquita, B., Philippot, P., & Boca, S. (1991). Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition and Emotion*, *5*, 435-465.
- Rosenthal, R. (1982). Conducting judgement studies. In K.R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 287-361). Cambridge: Cambridge University Press.
- Sadoski, M., Goetz, E.T., & Kangiser, S. (1988). Imagination in story response: Relationships between imagery, affect, and structural importance. *Reading Research Quarterly*, *23*, 320-336.
- Sadoski, M., & Quast, Z. (1990). Reader response and long-term recall for journalistic text: The roles of imagery, affect, and importance. *Reading Research Quarterly*, *25*, 256-272.
- Schank, R.C. (1979). Interestingness: Controlling inferences. *Artificial Intelligence*, *12*, 273-297.
- Silverstein, A., & Dienstbier, R.A. (1968). Rated pleasantness and association value of 101 english nouns. *Journal of Verbal Learning and Verbal Behavior*, *7*, 81-86.
- Stein, N.L. (1982). The definition of a story. *Journal of pragmatics*, *6*, 487-507.
- Stein, N.L. (1983). What's the point? *The Behavioral and Brain Science*, *4*, 611.
- Wainer, H., & Berg, W. (1972). The dimensions of de Maupassant: A multidimensional analysis of students' perception of literature. *American Educational Research Journal*, *9*, 485-491.
- Waltz, D.L., & Dorfman, M.H. (1983). The holes in points. *The Behavioral and Brain Science*, *4*, 612-613.
- Whissell, C.M., & Dewson, M. (1986). A dictionary of affect in language: III. Analysis of two biblical and two secular passages. *Perceptual and Motor Skills*, *62*, 127-132.

16 Final version in *Cognition and Emotion*, 7, 21-36

Wilensky, R. (1983). Story grammars versus story points. *The Behavioral and Brain Sciences*, 6, 579-623.