

## Chapter VI

### Evaluation

MEXICA is a *computer model* of a writing process. As stated in Chapter I and following Eglén (Eglén 1992, p. 2) computer models have three main purposes:

1. To verify that the theory they represent works (at least at a computational level).
2. To force the modeller to think about all the details of a hypothesis rather than just concentrating on the cardinal components of it.
3. To allow easy testing of the hypothesis under different circumstances, which in some cases can be difficult or not feasible to perform in human subjects.

This chapter evaluates MEXICA from these three perspectives. That is, since MEXICA is a *computer model* of a cognitive process, this chapter evaluates if the goals stated for a computer model have been achieved.

The chapter is organised as follows. Section 6.1 examines those aspects related to the evaluation of the theory. In order to verify that the engagement-reflection theory works, it is necessary to evaluate if MEXICA satisfies the main postulates of the model: the production of material without the use of explicit goals or predefined story-structures, and the production of stories as result of the interaction between engagement and reflection. To assess the latter aspect, this section introduces the engagement-reflection maps. These maps visualise the interaction between states during the writing process. The second step to verify if the theory works consists in assessing MEXICA's outputs. Four stories produced under different operation modes and with different values for the variable Tension to the Reader are evaluated based on criteria previously established. To complement this evaluation and compare MEXICA's outputs against other program's stories, fifty subjects completed a questionnaire to evaluate computerised stories. This questionnaire included four of MEXICA's stories and two produced by other programs.

The development of the computer model based on the engagement-reflection account forced the designer of MEXICA to think of all the details of the model. Particularly important are the creative process and the way of producing interesting stories. Section 6.2 analyses these two processes. The creative process is assessed by comparing the novelty of stories created by the system against the tales in the Previous Stories. The Associative Structure and the ACAS-process are examined to show how they contribute to generate novel knowledge in the system. In the same way, the method employed to produce interesting stories is analysed. This method is based on comparisons between the Tension to the Reader in the story in progress and the Previous Stories. This section includes, when it is considered appropriate, comparisons with previous works.

Section 6.3 evaluates if MEXICA offers enough facilities to test the model under different circumstances. The different features included in the system to allow easy testing under different contexts (e.g. the

definition of the Primitive Actions and Previous Stories, the different parameters defined by the user, and the utility of the reports produced by the system) are analysed.

## 6.1 Evaluating the theory.

As a computer model, MEXICA must be able to provide the information necessary to evaluate if the theory it embodies works. That is, it must demonstrate or refute the possibility of developing stories (at least a computational level) as a result of an engagement-reflection cycle. How can MEXICA demonstrate this? First, it is necessary to assess if MEXICA satisfies the main postulates of the engagement-reflection model and then evaluate if MEXICA's outputs are satisfactory.

During this work, two main premises have been mentioned as the key elements that distinguish this research from previous models (e.g. see Chapter I or Section 3.2.6):

1. During engagement MEXICA generates material without using explicit goals or predefined story structures.
2. MEXICA's outputs are the result of an engagement-reflection *interaction*.

Thus, it is necessary to demonstrate that these postulates are satisfied by the system. In the same way, this research claims that MEXICA is able to produce adequate outputs. Thus, it is necessary to evaluate if:

3. MEXICA's outputs can be considered as (frameworks for) stories.

That is, MEXICA must be able to produce texts that can be considered as short stories or at least as frameworks for short stories. Otherwise, the system cannot be considered as a model of the writing process. Although during this work MEXICA's outputs have been referred as stories, this author believes that computerised story telling is far from producing stories similar to those created by human beings. To make clear this distinction, in the questionnaire presented in Section 6.1.4 the outputs produced by the programs are referred as frameworks for short stories.

### 6.1.1 Production of Material during Engagement.

The main objective of the Engaged State is to produce material without using explicit goals or predefined structures. That is, to simulate part of the process where writers use writing as a way to discover what to say. Since MEXICA lacks any routine that employs goals or story structures to guide the development of a story during engagement, it is possible to conclude that all material produced during the Engaged State satisfies this requirement.

This objective is achieved through the use of the Abstract Representation and the routines to retrieve possible next actions from long-term memory. They allow MEXICA to recognise similarities between the context in the story in progress and analogous situations that have occurred in Previous Stories, and use them to select the next event in the tale.

By contrast, previous works require an explicit guide that indicates the path to follow to avoid illogical outcomes during story development. As a consequence these systems suffer some problems of rigidity. For

example, MINSTREL (Turner 1993) and GESTER (Pemberton, 1989) make use of explicit information about story structures to assure the production of coherent and interesting episodes. TALE-SPIN (Meehan, 1981) utilises goal-directed reasoning or inference functions to generate sequences of predefined actions to follow.

*In this way, MEXICA shows that a different method to predefined story structures and problem-solving techniques can be used to generate coherent sequences of actions during the story development.*

### 6.1.2 Engagement-Reflection Maps.

In previous chapters the routines that represent the Engaged and Reflective States have been analysed in detail and their main characteristics highlighted. However, to represent engagement and reflection in the computer model and generate texts is not enough in order to conclude that the theory works. One of the main claims of the theory is that outputs are the result of the *interaction* between engagement and reflection. To evaluate the degree of this interaction a set of tables called engagement-reflection maps are employed. They visualise MEXICA's dynamics in terms of engagement and reflection during the development of a story. The following is a map of the story *The princess who cured the Jaguar Knight* (see Section 5.1) in terms of the Engaged and Reflective States:

Reflective State	4-6,15
Initial Action	0
Engaged State	1
Reflective State	7, 16
Engaged State	2,3,8-11
Reflective State	13,17
Engaged State	12
Reflective State	14

Reflection 50.0%  
Engagement 44.4%

In order to explain how to interpret the map it is necessary to remember the way MEXICA generates this story. The user gives an initial event; then MEXICA generates three events under engagement and switches to reflection. That is, the initial action is given by the user at time=0, the first action during engagement is generated at time=1, the second at time=2, and the third at time=3. This can be represented in an engagement-reflection map as follows:

Initial Action	0
Engaged State	1-3

Map 1: Reflection 0.0%,  
Engagement 75.0%

The cell in black indicates the initial action (given by the user). Cells in white indicate the actions generated during engagement; cells in grey indicate the actions generated during reflection. The number of the map and the percentage of events generated during reflection and engagement is indicated at the bottom of the table. The numbers in the second column in the map indicate the time in which the actions were generated. The position of the entry in the table indicates the position of events within the story, with the start of the story at the top of the table and the end of the story at the bottom of the table. The order in which actions are presented in one cell is the same order they appear in the story. Thus, map 1 indicates that the order of the actions in the story produced so far is 0,1,2,3, end.

During reflection MEXICA inserts four actions at times 4, 5, 6 and 7 to satisfy the preconditions of the story in progress. Three of those actions are inserted at the beginning of the story and the fourth just after the action generated at time=1. This is represented as follows:

Reflective State	4-6
<b>Initial Action</b>	<b>0</b>
Engaged State	1
Reflective State	7
Engaged State	2,3

Map 2: Reflection 50.0%  
Engagement 37.5%

So far, this map indicates that the first three actions that the reader finds in the story in progress are generated during reflection at time=4, time=5 and time=6. The fourth event in the story is given by the user at time=0. The fifth event that the reader finds in the story in progress is generated during engagement at time=1, and so on.

Now MEXICA switches back to engagement and generates three events at time=8, time=9 and time=10. It switches to reflection but all the preconditions are satisfied, so no event is inserted. It switches again to engagement and produces an action at time=11, at time=12 and then an impasse is declared. At this moment the map looks as follows:

Reflective State	4-6
<b>Initial Action</b>	<b>0</b>
Engaged State	1
Reflective State	7
Engaged State	2,3,8-12

Map 3: Reflection 30.8%  
Engagement 61.5%

The only difference between this map and the previous one is that five events have been included during engagement to the story in progress between time=8 and time=12. MEXICA switches to reflection and

inserts an action at time=13 to satisfy the preconditions of the action generated at time=12. An event is inserted at the end of the story at time=14 to try to break the impasse. The map at this point looks as follows:

Reflective State	4-6
<b>Initial Action</b>	<b>0</b>
Engaged State	1
Reflective State	7
Engaged State	2,3,8-11
Reflective State	13
Engaged State	12
Reflective State	14

Map 4: Reflection 40.0%  
Engagement 53.3%

The impasse cannot be broken and the story is abandoned. MEXICA switches to reflection to analyse the story and inserts three events at time=15, time=16 and time=17 to produce the final version. So, the map looks as follows:

Reflective State	4-6,15
<b>Initial Action</b>	<b>0</b>
Engaged State	1
Reflective State	7, 16
Engaged State	2,3,8-11
Reflective State	13,17
Engaged State	12
Reflective State	14

Map 5: Reflection 50.0%  
Engagement 44.4%

This engagement-reflection map clearly shows:

- How the order in which the story is presented to the reader is not the same order in which it is generated.
- Engagement and reflection both contribute to writing different parts of the story in a dynamic manner. Furthermore, the story develops in a non-linear way rather than linearly progressing from the start of the story to its end.
- Most importantly, it clearly shows how the final story is the result of interactions between engagement and reflection: 50.0% of the story is created during reflection and 44.4% during engagement (the user gives the first action, i.e. 5.6% of the story). Notice how, as a result of the engagement-reflection

cycle, MEXICA intercalates different events during the development of the story. Each time a new action is inserted the story context is modified affecting the possible following actions to be retrieved. In the same way, all those events retrieved from memory whose preconditions are not satisfied provoke new actions to be included in the story.

The following table compares the percentage of events produced during engagement and reflection for each of the maps presented in this example:

MAP	Reflection	Engagement
1	0.0%	75.0%
2	50.0%	37.5%
3	30.8%	61.5%
4	40.0%	53.3%
5	50.0%	44.4%

The cells in dark grey highlight those states that generate more events. In this table we can observe how during the developing of the story there is a continuum switching in the leadership of the production of material between states. For example, in map 3 one can observe how the actions generated during engagement duplicate the number of actions generated during reflection. However, this trend changes through the development of the story and in the final version of the story (map 5) the number of actions generated in each state is very similar. The relation between actions generated during engagement and reflection changes due to the interaction between states and the different circumstances surrounding the developing of the story (e.g. how many actions with unsatisfied preconditions are retrieved). Such circumstances change from one story to another. For example, the following table shows the map of the story *The lovers* (see Section 6.1.3):

Reflective State	4-6
Initial Action	0
Engaged State	1-3
Reflective State	10,14,15
Engaged State	7
Reflective State	16
Engaged State	8,9,11,12
Reflective State	17
Engaged State	13

Reflection 44.4%  
Engagement 50%

Comparing this map with the one of the story *The princess who cured the Jaguar Knight* shows some of the differences in the dynamic of the system during the development of stories. Both stories are formed by

eighteen actions. However, during the development of the story *The princess who cured the Jaguar Knight* MEXICA retrieves more actions from memory whose preconditions are already satisfied than during the writing of *The lovers*. This situation can be observed in the princess' story map: see how the second engagement block has six actions, where four of them in a row (from 8 to 11) do not require to be justified (the other story only links three actions in a row). On the other hand, in the story *The lovers* more events are produced during engagement than in the story *The princess who cured the Jaguar Knight* (see the percentages at the bottom of each table). These differences only show how each story has its own dynamic during the development process.

*The material generated by MEXICA is the result of the interaction between engagement and reflection, and such interaction depends on the characteristics of each story in progress.*

### 6.1.3 Evaluating the Stories.

Evaluating MEXICA's output —or any other computerised storyteller— is problematic; a story that someone judges odd can be acceptable to another person. Thus, it is necessary to establish clear criteria to evaluate such outputs. A common method used from primary schools (see Enciclopedia Técnica de la Educación, 1990) to university literature courses to study fiction consists in separating and analysing the different properties or parts that form a particular piece of fiction. Notice that to isolate the elements that forms a novel, short story, etc., only works from a didactic point of view. That is, the separation between theme and style and narrative order is artificial and only acceptable for expositional and analytical reasons. It never happens in reality because what a novel says is inseparable from the way it is said (Vargas Llosa 1997, p.34)

What elements in the short stories produced by MEXICA must be evaluated? At the beginning of this research it is stated that MEXICA can produce stories that flow in a coherent way and are interesting (i.e. produce suspense, tension in the reader). Thus, these are two aspects that require to be assessed. One of the main differences between MEXICA and previous works is that MEXICA avoids using predefined story structures to produce a tale. Ensuing from this, the structure of MEXICA's outputs also needs considering. Another important difference between MEXICA and previous works has to do with the way the content of a story (i.e. what the story is about, what it says) is generated (MEXICA avoids using explicit goals or predefined story-structures). Thus, it is necessary to evaluate the content of MEXICA's stories. Other aspects like kind of language used, vocabulary, variety in the use of phrases, etc. are not part of this research (e.g. Natural Language Processing is out of the scope of this work) and therefore are not considered for evaluation. Thus, the stories produced by MEXICA will be evaluated according to the following four criteria:

- ❖ Flow and coherence
- ❖ Structure

- ❖ Content
- ❖ Suspense

The Collins English Dictionary defines coherence as “logical or natural connection or consistency.” Thus, this section analyses if the sequence of actions in MEXICA’s outputs flow in a logical or natural way. To produce interesting stories MEXICA builds suspense through the Tensional Representation method. Brewer and Lichtenstein (1982, p.481) affirm that a story involving suspense must include an initial event—which leads to significant positive or negative consequences for one of the characters in the narrative—and an outcome or resolution for such an event. Suspense is produced when the reader gets interested about the possible consequences that the initial event might have for the character. It ends when the outcome of the initial action is introduced into the narrative. In order to build up the suspense additional events are included between the initial action and its resolution.

Thus, this section studies if MEXICA’s stories produce suspense. The content has to do with what the story is about, what it says. Thus, this section examines if MEXICA’s stories have an adequate content and how complex it is. Finally, Clayton (1996) defines narrative structure as “how a writer orders the events of a story”(p.12). The order in which events are organised has a major impact on the story; e.g. the organisation of a murder fiction is not the same as the organisation of a fable. Some people propose that the structure of stories account for their enjoyment (Brewer and Lichtenstein 1982). Thus, this section studies if the MEXICA’s outputs have a clear structure. In this way, a good story would be that which:

- Flows in a coherent way.
- Has a clear structure.
- Has an adequate content.
- Is interesting for the reader.

These aspects are examined by this author in four stories produced by MEXICA under different operation modes and with different tensions to the reader. Two of MEXICA’s stories were generated under ER2 operation mode and two under E2 operation mode. To include another variable in the analysis, one of the ER2 stories and one of the E2 stories were forced to have low values for the variable Tension to the Reader, i.e. MEXICA was forced to produce two “boring” stories. In this way, it is possible to evaluate if the method employed to generate interesting stories works. Thus, the four stories can be classified as ER2, ER2-LTR (where LTR stands for Low Tension to the Reader, i.e. boring stories), E2, E2-LTR. These stories were selected because they were representative examples of the kind of stories MEXICA produced under these constraints.

#### **A. Analysis of the Stories.**

E2-LTR.

Although at the beginning Princess did not want to admit

it, Princess fell in love with Tlatoani. For long time Tlatoani and Lady had been flirting. Now, openly they accepted the mutual attraction they felt for each other. Princess hated Lady

The **E2-LTR** narrative is logical (the princess is in love with the Tlatoani; however, the Tlatoani and the lady like each other and as a consequence the princess hates the lady). However, the narrative is too short that it is difficult to find a clear story-structure (it gives the impression of being an extract from other story), it does not say anything (just that the princess hates lady) and fails to produce any suspense. The lack of these elements makes difficult to classify it as story.

#### E2

Hunter was an ambitious person and wanted to be rich and powerful. So, Hunter kidnapped Princess and went to Chapultepec forest. Hunter's plan was to ask for an important amount of cacauatl (cacao beans) and quetzalli (quetzal) feathers to liberate Princess. Farmer thoroughly observed Hunter. Then, Farmer took a dagger, jumped towards Hunter and attacked Hunter. Suddenly, Farmer and Hunter were involved in a violent fight. Hunter went in search of some medical plants and cured Princess. As a result Princess was very grateful to Hunter. Hunter and Princess went to the Great Tenochtitlan city.

The **E2** narrative presents a structure that allows distinguishing three main parts: the princess' kidnapping, the fight between the farmer and the hunter, and the hunter curing the princess and going back home. However, there is a lack of a coherent flow. In the first part of the story the farmer is observing the hunter; however, it is never explained how or when the farmer found the hunter (the forest is not an obvious place to find a farmer). Also, there is not an explanation of why the farmer suddenly attacks the hunter (one might assume that the farmer is trying to rescue the princess, but the situation is not clear). But the main logical gaps are in the following part. The hunter and the farmer are involved in a violent fight and in the next event, out of the blue, the hunter is searching medical plants to cure the princess. How and when does the fight stop? Does the hunter kill the farmer? How and when does the princess get injured? And finally, if hunter's intention was to get some economical benefit from the princess, Why does he go back home with her? The system is trying to introduce some suspense through the kidnapping, the fight, and the illness of the princess, but the lack of coherence prevents the reader of getting involved with the characters and the suspense is not produced. The same occurs with the content, i.e. the logical gaps avoids a proper linking of ideas and as a result there is not a clear notion of what the story is about.

#### E2-LTR

Jaguar\_knight was an inhabitant of the Great Tenochtitlan. Princess was an inhabitant of the Great Tenochtitlan. Jaguar\_knight was walking when Ehecatl (god of the wind) blew and an old tree collapsed injuring badly Jaguar\_knight. Princess went in search of some medical

plants and cured Jaguar\_knight. As a result Jaguar\_knight was very grateful to Princess. Jaguar\_knight rewarded Princess with some cacauatl (cacao beans) and quetzalli (quetzal) feathers.

The events in the **ER2-LTR** story flow in a coherent way and they are well structured. There is a clear introduction, an event that leads to negative consequences for the knight (degradation process) and a resolution. All the events are very logical: due to the wind there is an accident, the princess cures the knight and as a consequence he rewards her. However, the story is very boring. Nothing really happens since the tension produced by the accident is solved immediately. That is, any possibility of suspense is killed immediately. As a consequence of this situation, the content of the story is very simple: someone gets ill, his partner cures him and everybody is happy.

#### ER2 (The Lovers)

Jaguar\_knight was an inhabitant of the great Tenochtitlan. Princess was an inhabitant of the great Tenochtitlan. From the first day they met, Princess felt a special affection for Jaguar\_knight. Although at the beginning Princess did not want to admit it, Princess fell in love with Jaguar\_knight. Princess respected and admired Artist because Artist's heroic and intrepid behaviour during the last Flowery-war. For long time Jaguar\_knight and Princess had been flirting. Now, openly they accepted the mutual attraction they felt for each other. Jaguar\_knight was an ambitious person and wanted to be rich and powerful. So, Jaguar\_knight kidnapped Artist and went to Chapultepec forest. Jaguar\_knight's plan was to ask for an important amount of cacauatl (cacao beans) and quetzalli (quetzal) feathers to liberate Artist. Princess had ambivalent thoughts towards Jaguar\_knight. On one hand princess had strong feelings towards Jaguar\_knight but on the other hand Princess abominated what Jaguar\_knight did. Suddenly, the day turned into night and after seconds the sun shone again. Princess was scared. The Shaman explained to Princess that Tonatiuh (the divinity representing the sun) was demanding Princess to rescue Artist and punish the criminal. Otherwise Princess's family would die. Early in the Morning Princess went to Chapultepec forest. Princess thoroughly observed Jaguar\_knight. Then, Princess took a dagger, jumped towards Jaguar\_knight and attacked Jaguar\_knight. Jaguar\_knight was shocked by Princess's actions and for some seconds Jaguar\_knight did not know what to do. Suddenly, Princess and Jaguar\_knight were involved in a violent fight. In a fast movement, Jaguar\_knight wounded Princess. An intense haemorrhage arose which weakened Princess. Jaguar\_knight felt panic and ran away. Thus, while Tlahuizcalpantecuhtli (the god who affected people's fate with his lance) observed, Princess cut the rope which bound Artist. Finally, Artist was free again! Princess was emotionally affected and was not sure if what Princess did was right. Princess was really confused. The injuries that Princess received were very serious. So, while praying to Mictlantecuhtli (the lord of the land of the dead) Princess died.

**ER2** story presents a clear structure. At the beginning the main characters are introduced in the story and the emotional links that tie them are clearly established. Then, the main degradation process starts: the knight kidnaps the artist. This situation not only affects the artist but also reaches the princess who is confronted with two options: following the man she loves or fighting against what she thinks is abominating. A sign sent by the gods (the eclipse) makes her mind up; so, she decides to go and rescue the artist. The rescue scene includes the beginning of a new degradation process: her lover injures the princess. Finally the resolution arrives: the artist is liberated, the princess dies questioning her own behaviour and the knight escapes. The whole story flows in a very coherent way, i.e. each action performed in the tale is justified. Particularly important is the event where divine forces command the princess to go and rescue the artist: it explains princess' behaviour and drives the story towards its tragic end. The content is also interesting: the story talks about someone at the crossroads, the dilemma of deciding between two opposite

options and the conflict it produces. The fact that her lover is who kills the princess produces an ironic situation and suggests how, at the end, any decision would have brought a sad end. The dilemma, the fighting and wounding events produces suspense in the reader and makes an interesting story.

From this analysis it is concluded that the story ER2 (also referred as *The Lovers*) better meets criteria of coherence, structure, content and suspense established at the beginning of this section as a characteristics of a good story. The ER2-LTR story satisfies the coherence and structure points; the content is not very well managed and there is a lack of suspense. Thus, this story is classified as the second best. The E2 story presents a not very clear structure and a serious problem with the coherence that produces a lack of content and problems with the suspense. So, this story is classified as third. The E2-LTR story is the worst since it does not satisfy any aspect.

	ER2	ER2-LTR	E2	E2-LTR
Coherence	Y	Y	N	N
Structure	Y	Y	~	N
Content	Y	~	N	N
Suspense	Y	N	~	N

(Y= yes, N= no, ~= maybe)

This initial analysis shows differences between the four types of stories. These differences are more rigorously examined in the form of a questionnaire in the following section.

#### **6.1.4 The Questionnaire.**

In order to test the initial results obtained after the analysis in Section 6.1.3 and compare MEXICA's outputs with other programs' stories, a questionnaire to find people's opinion about a group of computerised stories was designed. Its purpose was to ask a heterogeneous group of persons to assess the flow and coherence, structure, content, suspense and overall quality of short stories produced by MEXICA and other computer programs. The full questionnaire is given in Appendix G.

##### **A. Description of the Questionnaire.**

The questionnaire was presented as a PhD research in computer-based story generation. It collected some personal data of the participants: their age, sex, nationality and last academic grade obtained. Seven stories were included: four produced by MEXICA (the stories analysed in Section 6.1.3), one by GESTER, one by MINSTREL and one by the author of this research. These stories were selected by the author as representative examples of the kind of stories these programs produced. For example, in the case of GESTER's story, Pemberton included it as a part of an international paper published by her (Pemberton 1989). In the case of MINSTREL's story, it is the first tale the reader finds in Turner's PhD thesis (Turner

1993). The author of this work created the last story included in the questionnaire; this fact is not mentioned to the subjects participating to avoid prejudices in their answers. This story was written trying to imitate the same language used in the other stories in the questionnaire. However, this author made an effort to produce the best structure, content, coherence and suspense he could in the story. This story, called the HUMAN story, provides a comparison between human and computer-generated stories. The presentation order of the seven stories in the questionnaire was selected at random. This is the result:

- Story #1 MEXICA E2
- Story #2 GESTER
- Story #3 MEXICA ER2-LTR.
- Story #4 HUMAN story.
- Story #5 MEXICA E2-LTR.
- Story #6 MINSTREL
- Story #7 MEXICA ER2

Subjects were asked to evaluate the flow and coherence, structure, content, suspense and overall quality of each of the seven narratives. The questionnaire did not include any definition or explanation of what is understood for each of these five aspects. Since it was required that all subjects participating in the questionnaire must have finished their A-levels (or equivalent), it is assumed that (at least) they possess a basic understanding of these terms. For each aspect to be evaluated there were five possible answers ranging from “very good” to “very poor”. An additional question to know how much subjects liked each story was included. At the end of the questionnaire subjects ranked the seven stories: one for the best, seven for the worst. Also, a space was provided to write any comments they felt like expressing about the questionnaire.

## **B. Description of Subjects.**

Fifty subjects answered the questionnaire; they were required to be over 21 and have passed A-level (or equivalent) grade. This constraint produced a subject pool (42% female, 58% male). The average age of the participants was 30.92; the youngest was 25 years old, the oldest 62. The percentage of participants classified by nationalities and their grade of education is showed in table 6.1: 38% of them were Americans (Brazilians, Canadians, Mexicans and USAns) and 62% Europeans (Austrians, British, Dutch, French, Germans, Greeks, Norwegians and Polish). Regarding their education, 92% of them were graduated, 68% possessed postgraduate degrees and 6% did not answer the question.

<b>Nationality</b>											
Mexican	British	German	Brazilian	Polish	Norwegian	Greek	USA	French	Canadian	Dutch	Austrian
20%	20%	16%	12%	10%	6%	4%	4%	2%	2%	2%	2%

Education					
Unknown	Diploma	Bachelor	Master	MPhil	PhD
6%	2%	24%	46%	4%	18%

**Table 6.1 Nationalities and educational level of the subjects that answered the questionnaire.**

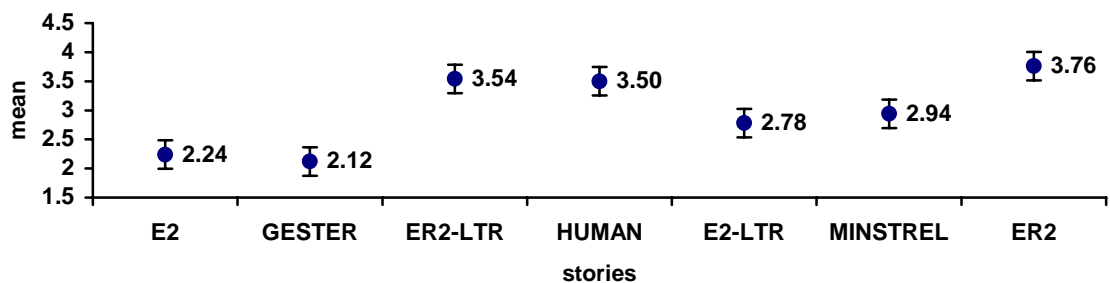
Although English was a second language for most of the subjects, stories were short and simple enough for the language not to be an issue. All questionnaires were sent and returned via electronic mail.

### C. Results.

In all cases the Friedman analysis of variance is applied to obtain the level of significance. Following Greene and d'Oliveira (1981) values from 1 to 5 are allotted to the five-point scale between very poor and very good (and between not at all and a lot). For each evaluated aspect, the assessment of each story is equal to its mean. The confidence interval at 95% level is included for each result.

**Narrative flow and coherence.** The following are the results obtained from the evaluation of the narrative flow and coherence ( $p < .01$ ): MEXICA ER2  $3.76 \pm 0.29$ , MEXICA ER2-LTR  $3.54 \pm 0.26$ , Human  $3.50 \pm 0.35$ , MINSTREL  $2.94 \pm 0.31$ , MEXICA E2-LTR  $2.78 \pm 0.27$ , MEXICA E2  $2.24 \pm 0.26$  and GESTER  $2.12 \pm 0.28$ .

**Fig. 6.1 Evaluation of the Narrative Flow and Coherence**



The rating of MEXICA's stories meets the results of the preliminary analysis of Section 6.1.3. The following comments made by two subjects, mention the lack of coherence in story E2 (i.e. story #1 in the questionnaire):

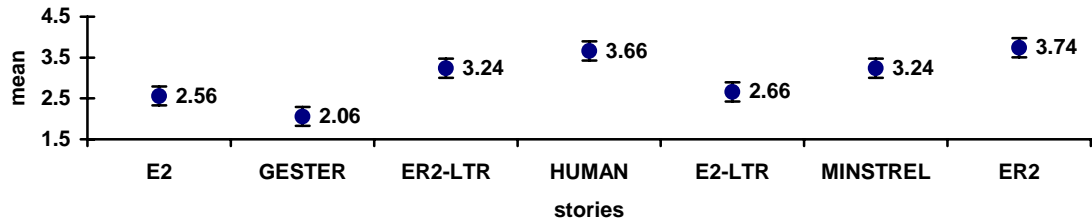
#1 comment: there's a problem. We never hear why princess needs to be cured -- has she been injured? why and when? ...

Narrative #1 contains an illogical jump; it looks like a part is missing from an otherwise good story.

The two stories produced under ER2 operation mode (stories #3 and #7 in the questionnaire) obtained — between the seven stories— the highest score in the evaluation (together with the Human story!). These results suggest that MEXICA's ER2 stories have a satisfactory flow and coherence.

**Structure.** The following are the results obtained from the evaluation of the narrative structure ( $p < .01$ ): MEXICA ER2  $3.74 \pm 0.24$ , Human  $3.66 \pm 0.29$ , MINSTREL  $3.24 \pm 0.27$ , MEXICA ER2-LTR  $3.24 \pm 0.27$ , MEXICA E2-LTR  $2.66 \pm 0.28$ , MEXICA E2  $2.56 \pm 0.27$  and GESTER  $2.06 \pm 0.30$ .

**Fig. 6.2 Evaluation of the Narrative Structure**



In the case of MEXICA's outputs —as expected— the two stories generated under ER2 operation mode obtained a higher evaluation than those generated under E2 operation mode. However, contrary to the analysis done in Section 6.1.3, the majority of the subjects considered that the E2-LTR story had a better structure than the E2 story. Nevertheless, the confidence level indicates no difference.

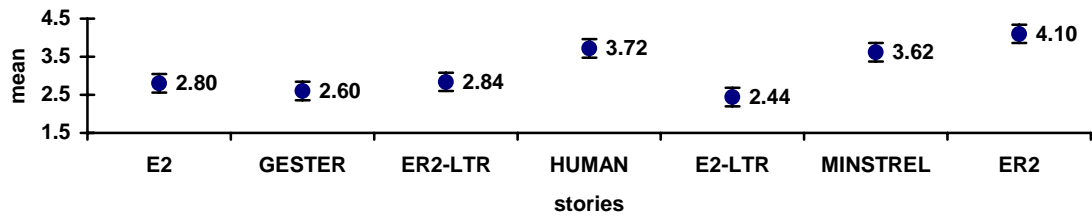
Although MINSTREL and GESTER make use of explicit story structures information to build their stories, MEXICA got the highest score in this aspect. This result suggests that it is possible to produce stories with an adequate structure without explicitly predefining it. Notice that MEXICA ER2 story obtained a higher evaluation than the human story. This result seems to suggest that, in opinion of the subjects participating in the questionnaire, MEXICA is able to structure a computer-style story better (or at least with the same quality) than the author of this work is.

**Content.** The following are the results obtained from the evaluation of the content ( $p < .01$ ) (see Fig. 6.3): MEXICA ER2  $4.10 \pm 0.26$ , Human  $3.72 \pm 0.26$ , MINSTREL  $3.62 \pm 0.25$ , MEXICA ER2-LTR  $2.84 \pm 0.24$ , MEXICA E2  $2.80 \pm 0.27$ , GESTER  $2.60 \pm 0.28$  and MEXICA E2-LTR  $2.44 \pm 0.27$ .

Regarding MEXICA's outputs, this time the results obtained from the questionnaire supported the analysis in Section 6.1.3: ER2 got the highest score, followed by ER2-LTR, E2 and E2-LTR. Fig. 6.3 clearly shows an important difference between ER2 and the rest of MEXICA's stories.

MEXICA obtained the highest and lowest evaluation in this aspect between all the stories. That is, MEXICA produced the story with the best content and with the worst content. These results reflect the importance of the suspense and reflective processes in the production of satisfactory contents.

Fig. 6.3 Evaluation of the Content



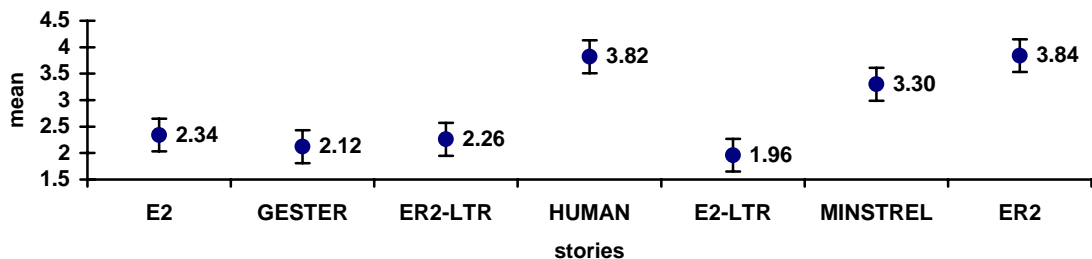
Subjects evaluated the content of the ER2 story higher than the content of the human story. This result took the attention of this author since the human story included some humour, a characteristic not found in any of the other six stories. However, although this fact was important for some of the subjects, it did not influence the overall result. The following comments show this situation:

4 was a bit humorous, which was pleasant.

Number 7 was good because of the rise and fall of suspense, and was only beaten by number 4 because of the humour in narrative number 4.

**Suspense.** The following are the results obtained from the evaluation of the suspense ( $p < .01$ ): MEXICA ER2  $3.84 \pm 0.30$ , Human  $3.82 \pm 0.28$ , MINSTREL  $3.30 \pm 0.29$ , MEXICA E2  $2.34 \pm 0.26$ , MEXICA ER2-LTR  $2.26 \pm 0.25$ , GESTER  $2.12 \pm 0.30$  and MEXICA E2-LTR  $1.96 \pm 0.26$ .

Fig. 6.4 Evaluation of the Suspense



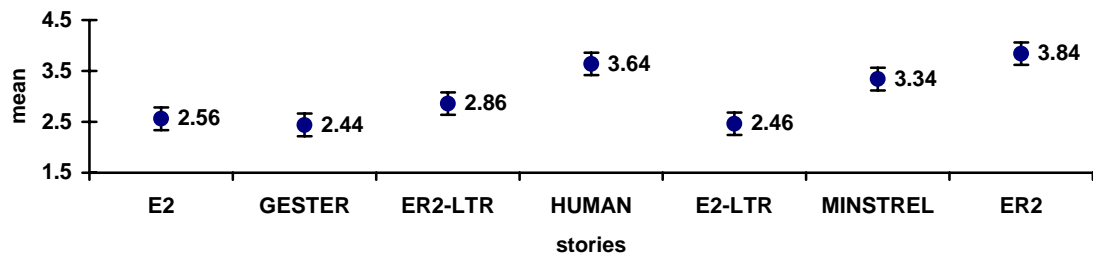
With respect to MEXICA's outputs, the results obtained meet what was expected: ER2 got the highest score followed by E2 story, while ER2-LTR story got the third place followed by E2-LTR story. That is, those stories forced to maintain low values for the Tension to the Reader got lower scores in the suspense assessment than those produced normally. Fig. 6.4 shows a difference between ER2 and the rest of MEXICA's stories, as it was mentioned in the initial analysis in Section 6.1.3. The following is a comment made by one of the subjects regarding ER2-LTR story (i.e. story #3 in the questionnaire):

Narrative 3 is coherent but everything goes normal, correct (except for the accident which is immediately solved) and unavoidably boring. There is no motivation (problem to be solved, suspense, drama, hate, love, crime, etc.) for a short story.

MEXICA ER2 story together with the human story got the highest score. These results suggest that the process followed by MEXICA to create suspense and interesting stories work adequately.

**Overall Quality.** The following are the results obtained from the evaluation of the overall quality ( $p < .01$ ): MEXICA ER2  $3.84 \pm 0.32$ , Human  $3.64 \pm 0.31$ , MINSTREL  $3.34 \pm 0.30$ , MEXICA ER2-LTR  $2.86 \pm 0.25$ , MEXICA E2  $2.56 \pm 0.28$ , MEXICA E2-LTR  $2.46 \pm 0.28$  and GESTER  $2.44 \pm 0.30$ .

**Fig. 6.5 Evaluation of the Overall Quality**

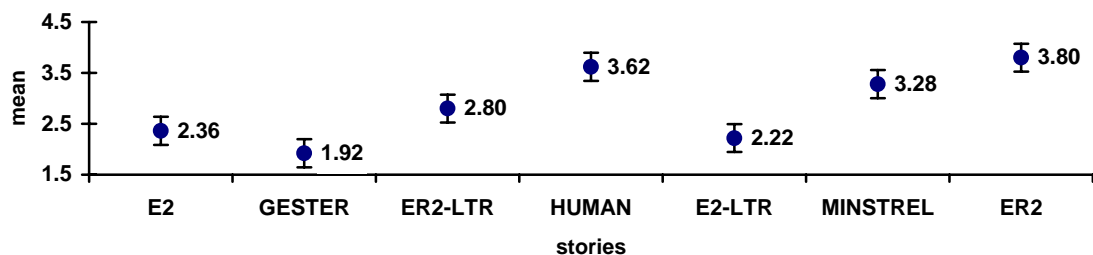


Regarding MEXICA's outputs, the results of the assessment of the overall quality meet the conclusions in Section 6.1.3: ER2 was considered the best story, followed by ER2-LTR, E2 and E2-LTR.

Subjects evaluated MEXICA-ER2 as the story with the best overall quality. These results suggest that MEXICA is able to produce satisfactory computerised short stories.

**Liking of the stories.** The following are the results obtained from the evaluation of the liking of the stories ( $p < .01$ ): MEXICA ER2  $3.80 \pm 0.28$ , Human  $3.62 \pm 0.25$ , MINSTREL  $3.28 \pm 0.25$ , MEXICA ER2-LTR  $2.80 \pm 0.21$ , MEXICA E2  $2.36 \pm 0.21$ , MEXICA E2-LTR  $2.22 \pm 0.25$  and GESTER  $1.92 \pm 0.29$ .

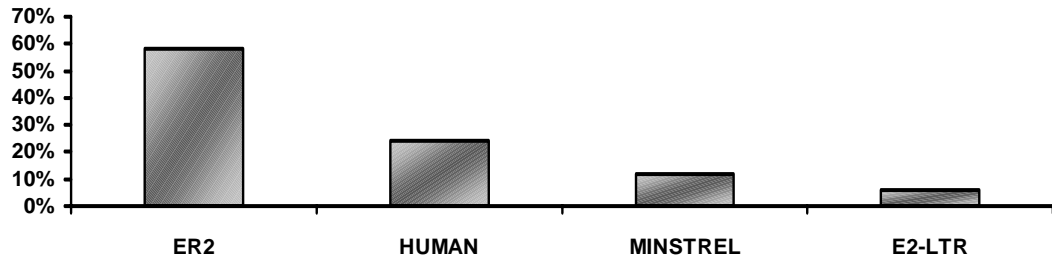
**Fig. 6.6 Evaluation of how much the Stories were Liked**



Subjects evaluated MEXICA ER2 story as the favourite story.

Figure 6.7 shows those stories ranked as the best: 58% of the subjects considered MEXICA ER2 as the best story, 24% chose the human story as the best one, 12% selected MINSTREL and 6% MEXICA E2-LTR story.

**Fig. 6.7 Evaluation of the Best Story.**



To establish a possible correlation between these results and the length of stories, the number of words in each narratives was registered: MEXICA ER2 was the longest story in the questionnaire (328 words in 35 lines); the Human was the second longest story in the questionnaire (221 words in 20 lines); MINSTREL was the third longest story (177 words in 20 lines). This analysis suggested that there was a correlation between those stories rated as the best and their lengths. However, it is interesting to point out that, although MEXICA E2-LTR was the shortest story in the questionnaire (41 words in 4 lines), it was considered as the best story by 6% of the subjects. Nobody considered MEXICA E2 (95 words in 10 lines), GESTER (179 words in 19 lines) or MEXICA ER2-LTR (66 words in 8 lines), which were longer, as the best story.

#### **D. Some Comments.**

MEXICA and MINSTREL are the programs that obtained the highest scores. Each row of the following table compares the comments made by the same subject about the stories produced by these programs.

<b>MINSTREL's story</b>	<b>MEXICA's ER2-story</b>
6 was hard to place. It was a good story, but it made many obvious statements about what was happening—statements that could be easily inferred.	7 was poorly written but was the most interesting. •
Narrative #6 has a good content but contains a couple of the irritating telegram phrases that narrative #2 is full of.	Narrative #7 is good. Only in the end it becomes a bit incoherent.
#6 Comment: story is suspenseful. There are linguistic problems: sentence 3 and 4 duplicate information which is confusing. Too much is spelled out throughout the story. More should be implicit. Let the readers do the inference! This is a cool story, I really liked it a lot. the ending is unexpected and funny!!!! Was it really machine generated or is this a filler???	#7 Anaphora resolution specially bad! The ending is a anti-climactic. The overall connection of the story and complexity is impressive, though. The princesses' conflict is well worked on and makes a good story.
Narratives 2,4,5 and 6 are OK. They can have some inconsistencies but overall are fine.	Narrative 7 is dense, there are different characters with ambiguous feelings and there is a story (action) really happening.

**Table 6.2 Comparison of comments about stories #6 and #7.**

These comments are included since they allow the reader to visualise how some subjects perceived the evaluated stories, information that cannot be inferred from the answers in the questionnaire. For example, to question if story #6 was really machine generated (see the third row in Table 6.2) is a good compliment to MINSTREL; the same occurs with this person's opinion regarding ER2-story's complexity.

### **E. Conclusions.**

- Stories produced under ER2 operation mode have better coherence, structure and content than those produced under E2 operation mode (people evaluated higher these aspects in the stories generated under ER2 than those produced under E2).
- The use of the Tension to the Reader as a method to produce interesting stories (suspense) works adequately. Subjects perceived the lack of suspense when outputs were forced to have low values of tension.
- Under the ER2 operation mode, MEXICA produces its best stories.
- Subjects evaluated MEXICA ER2 story higher than those representing other important systems.
- Subjects evaluated the coherence, structure, content and suspense of the MEXICA ER2 story higher than the ones of the story produced by the author of this work. This result suggests that subjects consider that MEXICA is able to produce computer-style stories (at least) similar to the ones produced by this author.

At the beginning of this section the criteria to evaluate stories was defined: coherence, structure, content and suspense. MEXICA ER2, MINSTREL and the human's stories got the highest scores in these aspects. And such stories were the ones people like the most and the ones that got the best evaluation in the overall quality as a framework for a short story point. This suggests that the selection of these aspects as criteria to evaluate stories is adequate.

Finally, it is interesting to notice how some of the subjects who answered the questionnaire seemed to enjoy it. For example, two subjects commented:

```
The stories were quite funny!
What do you want to show or find out?

It was really fun doing this experiment!!
```

However, others seemed to be against computerised story production:

```
I would suggest to stop the experiment with PCs literature, unless
the
purpose is to prove they are not the right poets.
```

*The results obtained from the questionnaire support the thesis that the engagement-reflection cycle is a viable model to produce adequate frameworks for short stories.*

## **6.2 Evaluating the Details of the Model.**

MEXICA is a good example of how a computer model forces the modeller to think about all the details of a theory. Sharples' (1996) account of the writing process is very general, i.e. it does not include details about the cognitive processes performed during writing. Thus, an important part of this research consists of specifying all those details in order to create and implement the computer model. The computer model topic of this research is very complex and includes many processes that interact continuously to produce an output. However, there are two that —due to their importance— require a special consideration: the creative process followed by MEXICA and the method employed to produce interesting stories.

### 6.2.1 Evaluating the Creative Process in MEXICA.

As mentioned in Section 3.2.6 and following Sharples (1996), creativity is defined here as the production of novel and adequate outputs. In Sections 6.1.3 and 6.1.4 the appropriateness of MEXICA's outputs has been assessed. In this section the capacity of MEXICA to produce novel stories is evaluated.

Story	Topic
Previous Story #1	<u>Love and Disloyalty</u> . This story is about a knight who falls in love with his brother's girlfriend and decides to kidnap her.
Previous Story #2	<u>Love and Obsession</u> . This story is about a princess who falls in love with a knight. When she realises that he is in love with another woman, the princess decides to kill her rival.
Previous Story #3	<u>Envy</u> . This story is about a prince who envies his father position. The prince's ambition is so big that he abandons his father in the forest after he suffers an accident.
Previous Story #4	<u>Love and Obsession</u> . This story is about a knight who is in love with a woman who is attracted to a different man. The knight decides to attack his rival.
Previous Story #5	<u>Valour</u> . This story is about a kidnapped princess who is rescued by a farmer. The farmer is wounded during the rescue and dies.
Previous Story #6	<u>Gratitude</u> . This story is about a hunter who saves the life of a Tlatoani, and later the Tlatoani does the same for the hunter.
The princess who cured the Jaguar Knight	<u>Love and Revenge</u> . This story is about a kidnapped princess who is rescued by a knight. The princess falls in love with the knight. However, she realises that the knight murdered of her father and decides to kill him.
The lovers.	<u>Love and Values</u> . This story is about a princess who is in love with a man whose values clash with the princess' values. Thus, she has to decide between following her values or following her man.

**Table 6.3 Relation of the themes in the Previous Stories and two tales created by MEXICA.**

The criterion used to evaluate the novelty of MEXICA's outputs consists in comparing the topics of the tales produced by the system against the topics of the set of Previous Stories. That is, since all structures in

memory representing knowledge are formed from the set of Previous Stories given by the user, a novel story would be that whose topic is original with respect to any of the tales' topics in the Previous Stories. This method of assessment is inspired by Boden's ideas. For her, "The psychological sense [of creativity] concern ideas ... that are fundamentally novel with respect to *the individual mind* which had the idea." (Boden 1992, p.32) In the same way, in this research the computational sense of creativity concern stories produced by the system that are novel with respect to the set of Previous Stories loaded in its memory. Table 6.3 shows a relation of the topics in the file of Previous Stories and the topics of two stories created by MEXICA.

Although some events in the new tales might be similar to some episodes in the Previous Stories (Section 5.3.1 offers a clear explanation of why this similarity arises) Table 6.3 shows how the topics of the stories produce by MEXICA are novel with respect to the Previous Stories.

*Thus, MEXICA is capable of producing novel and adequate stories.*

#### **A. Processes Involved in the Production of Novel Stories.**

Although the whole system participates during the production of a novel tale, two core elements are responsible for bringing novel events into working memory: the Associative Structures and the ACAS-Process.

Associative Structures are dynamic organisations built from characters' contexts used to match atoms in long-term memory. Each time an Associative Structure is transformed, the requirements to match an atom are also modified and options not available in previous searches are now reachable. Although only two types of transformations have been implemented, other ways to modify the structure can be easily developed and included in the code.

The ACAS-Process allows matching atoms with a greater number of elements than the Associative Structure. That is, the ACAS-Process is analogous to transforming the Associative Structure by adding new Emotional Links or Tensions, in order to make it match a bigger atom. However, there is an important difference between the ACAS-Process and the transformation process. New elements added to the Associative Structure through a transformation routine are fixed; they cannot be removed until a new transformation is executed. By contrast, the ACAS-Process simulates some kind of "dynamic added elements". That is, through the ACAS-Process the Associative Structure behaves as if it had a fixed part (representing the core characteristics of the structure) and a dynamic part able to adapt to different atoms each time a match is tried. When an atom with similar characteristics but greater number of elements than the Associative Structure is found, the ACAS-Process produces the same effect as if all of the missing elements were added automatically to the Associative Structure in order to make it match the atom. Because this process can be repeated several times in a search, the overall effect is that of an Associative Structure able to adapt to various atoms with different characteristics.

## **B. Production of Novel Knowledge.**

MEXICA transforms Associative Structures or executes the ACAS-Process when contexts not represented in long-term memory arise, or when no atom can be matched after all options brought to Working Memory have been rejected by the filters.

As a result of these processes, during story development MEXICA is able to generate contexts not present before in long-term memory. For example, in Section 5.2—after the action ENEMY KIDNAPPED PRINCESS is executed— the resulting princess and jaguar knight’s contexts cannot be found in the Abstract Representation. Two new contexts are therefore generated by the system. This is indicated in the report by the expression **PRINCESS: 3,7 ASI-50 JAGUAR\_KNIGHT: 1,7 ASI-50** which means that the princess matched an atom 50% equal to the Associative Structure I and the knight matched and atom 50% equal to the Associative Structure I. That is, because new contexts have been created MEXICA cannot match an atom equal to the contexts in memory. Thus, the ACAS-Process is executed to match an atom 50% similar to the original context.

Other examples illustrating the same phenomena can be found in Section 5.2. The ability to produce contexts not present in the Abstract Representation is a very important characteristic of the system that allows producing novel stories. In other words, different authors see writing as a way of transforming experiences into new realities (see Section 2.2.1). Inspired by those ideas, MEXICA transforms information extracted from the Previous Stories into new contexts, resulting in the production of novel tales.

*MEXICA is able to create contexts not present in the system before.*

Of all the previous works studied in this thesis, only MINSTREL (Turner 1993) has explicit routines to evaluate the novelty of the story in progress and, when necessary, to modify events in memory to produced novel material not present before in the system. It is difficult to compare MINSTREL’s heuristics against MEXICA’s because they have very different approaches. In MINSTREL such heuristics (called TRAMs) attempt to create novel situations by transforming and adapting episodes in memory in order to instantiate predefined themes. By contrast, MEXICA attempts to recognise similar contexts to generate novel sequences of actions.

Nevertheless, some differences can be mentioned. An important and clear one is that MINSTREL is not able to produce novel themes to develop stories; all possible topics in MINSTREL are predefined. And furthermore, the structure that any story can have has also been decided in advance. In this way, creativity in MINSTREL means to generate novel episodes to instantiate pre-determinate structures. By contrast, in MEXICA the themes of stories are not defined a priori. Instead, they develop as the program is used repeatedly. Of course, any topic developed by MEXICA ultimately depends on the information stored in the Previous Stories. However, the flexibility of the processes included in the system permits developing themes not explicitly defined in the Previous Stories.

TRAMs can produce interesting and complex episodes. However, sometimes they look as if they were written to achieve particular events and able to work properly only under certain circumstances (see Section 2.4.3). If MINSTREL is fed with new themes and episodes it is necessary to write a new set of TRAMs that can transform them. By contrast, MEXICA's routines are made to identify similar contexts independently of the theme and so can adapt to new situations. Furthermore, parameters like the ACAS-Constant, the Novelty Percentage and Novelty Constants allow experimenting with different aspects of the creative process.

*Based on the information stored in the Previous Stories MEXICA is capable of producing novel tales. The Associative Structure and the ACAS-Process are the main routines used to retrieve original material from memory. MEXICA is able to produce novel structures in memory (representing new knowledge) that can be used to generate original tales.*

## **6.2.2 Evaluating the Tensional Representation as a Way to Produce Interesting Stories.**

Conflicts, and the tension they produce in the reader, are essential components to create interesting stories. And although the ACAS-Process and Associative Structures are useful to retrieve events, it is no guarantee that the actions retrieved from long-term memory will produce an interesting story.

MEXICA solves this problem through the use of Tensional Representations. They represent degradation-improvement processes that register the tension of the story in progress (see Figure 5.3). By default a story is classified as producing enough interest when its Tensional Representation is at least 50% equal to the Tensional Representation of any of the Previous Stories (also called frames). Thus, in order to evaluate the story in progress MEXICA compares its Tensional Representation against all available frames, selects the frame most similar to it, and checks if they are at least 50% equal (see example in Section 5.2).

Although the evaluation of the story in progress is based on the similarity with a frame, the process of developing a tale is flexible enough to allow forming new Tensional Representations instead of just copying old ones. To illustrate this situation the reader can compare in Section 5.2 the Tensional Representations of the story in progress and all the Previous Stories. The result of the evaluation of the Tensional Representation is used to set the Tensional Guidelines that influence the production of material during the Engaged State (see Section 5.3.7).

*MEXICA verifies if a story is interesting by comparing the Tensional Representation of the story in progress against the Tensional Representation of all Previous Stories. The method is flexible enough to allow forming new Tensional Representations instead of just copying the old ones.*

The effectiveness of this method has been tested in the questionnaire described in Section 6.1.4. As reported earlier, those stories with low Tension to the Reader were the worst evaluated in suspense, while the stories with high Tension to the Reader obtained the best assessments.

MINSTREL (Turner, 1993) has a different way to produce interesting stories. Since the writing process consists of instantiating predefined structures (known as themes), MINSTREL simply includes in those structures elements that will make the story interesting. For example, the theme “Hasty-Impulse-Regretted” (Turner, 1993 p.267) is inspired by Romeo and Juliet’s tragic end. When Romeo finds Juliet’s body, he kills himself thinking she is dead. So, the message of this theme is to avoid doing things in haste, especially when their consequences are irreversible. MINSTREL represents this theme as a set of structures that must be instantiated. These structures can be described as follows. Character X has a belief that prevents him from achieving a goal. This, and his hasty personality, makes him do something irreversible. Character X learns something new which supersedes the evidence for his earlier belief. So, now that Character X has a different belief he wants to retract from his earlier action but he cannot because it is irreversible (Turner 1993, p.268).

This example illustrates how the events that will produce tension in the reader and make the story interesting are defined in MINSTREL’s structures before the writing process starts. By contrast, in MEXICA those elements are introduced in the story as a part of the writing process. That is, MEXICA performs constant evaluations of the story in progress to try to avoid the production of boring tales. Filters also help to eliminate actions that do not contribute to develop interesting stories.

Something similar to MINSTREL occurs in GESTER (Pemberton 1989). The grammar assures that a story includes an initial situation, a development (which involves a complication, a plan to overcome the problem, etc.), and a final situation. TALE-SPIN (Meehan, 1981) does not include any process to evaluate the interest produced by a story.

### **6.3 Evaluation of MEXICA as a Research Tool.**

One of the main objectives of this research is to give the user the opportunity to test the system under diverse circumstances. This section examines if this objective is met.

The computer program topic of this research can be simplified to two main components: the structures in memory and the processes that use those structures to generate new stories. If users can control the content of such structures and manipulate the features of the writing processes easily, the goal of allowing easy testing under different circumstances is achieved.

All memory structures in MEXICA are built from the files of Primitive Actions and Previous Stories. Thus, this section analyses the facilities offered by the system to define both of them and mentions how such structures affect the writing process. In the same way, the parameters definable by the user offer a

straightforward method to control some aspects of the writing process. Thus, they are also studied. Finally, the utility of reports produced by the system is considered.

### **A. Primitive Actions.**

Primitive Actions, their preconditions and postconditions are used to build the Primitive Actions Structure. They are definable by the user through a simple language known as Primitive Actions Definition (PAD) (see Appendix A). PAD offers the following facilities:

- The use of the symbol star ('\*') to refer to multiple actors or types of Emotional Links.
- The symbol percentage ('%') to allow establishing the intensity of an Emotional Link between characters at run time (i.e. at the moment of developing a story).
- Linked Characters to allow establishing some of the actors participating in an Emotional Link at run time, etc.

Examples are shown in Section 4.1. These facilities permit easy definition of complex events involving multiple preconditions and postconditions. This characteristic contrasts with the Conceptual Dependency model used by TALE-SPIN (Meehan, 1981) and MINSTREL (Turner, 1991) where the definition of some events can involve very complicated structures. During this work it has been pointed out in different occasions the way preconditions affect the writing process, and how postconditions modify the story-world context and therefore the process of retrieving actions from memory. Thus, the opportunity to define them not only allows establishing control over the content of the structures in memory but also influencing the developing process (e.g. satisfaction of preconditions, search of equivalent actions to break an impasse).

*Thus, the possibility of defining and modifying actions, preconditions and postconditions in a very simple way gives the user the opportunity to control the content of the Primitive Actions Structures. The information in such structures constrains some of MEXICA's processes (e.g. satisfaction of preconditions).*

### **B. Definition of Previous Stories.**

The user defines all Previous Stories in the system through a language called Description of Previous Stories (DPS). DPS allows the specification of an unlimited number of stories without any restriction in their length. There are only two requirements to define them: the stories must exclusively include events defined as Primitive Actions and follow the DPS' syntax rules. The option of control over the content of the Previous Stories is important because the Concrete, Abstract and Tensional representations are created with information extracted from this file. In other words, MEXICA's structures representing knowledge are built from the Previous Stories. In this manner, the user can influence the way these structures are organised. For example, through the Previous Stories the user defines relations between story-contexts and logical possible next actions to perform. Thus, MEXICA knows that when a character is ill (story-context) it is necessary to look for a doctor (logical possible action). The control of the content in the Previous

Stories is also useful for other purposes. For instance, the example in Section 5.5.2 shows how the user can select certain Previous Stories to create the right conditions to test some of the characteristics in MEXICA.

*The Previous Stories determine the content of the main structures (representing knowledge) that guide the writing process in MEXICA. Thus, through their definition, the user can influence the way these structures are organised and therefore the writing process.*

### **C. Parameters Definable by the User.**

MEXICA offers the user the possibility of manipulating twenty-three different parameters that constrain the story development process (see Table 4.1). The number of experiments and tests that can be done by varying all the different parameters available in the system is so large that only some of them can be reported in this work. In the experiments described in this thesis the following parameters are manipulated: ACAS-Constant (Section 5.4.1), Forbidden Actors (Section 5.4.2), the Operation Modes (Sections 5.4.3 and 5.5), the Initial State (Section 5.5.1) and the CtEg-Rf (Section 5.5.2).

The flexibility that these parameters offer allows, for example, confirming some hypothesis developed at the beginning of this research (e.g. the differences reported between operation modes). Also, the testing under different conditions has raised aspects not considered earlier (e.g. the large impact that in the developing of a story can have the different ways of instantiating characters). This information can be very useful to develop novel computer programs that can explore new aspects of the Engagement and Reflection cycle. It is also hoped that it will help to improve or develop new models of the writing process.

*More than 20 parameters definable by the user allow constraining and modifying different features of the writing process in MEXICA.*

### **D. Reports in MEXICA.**

MEXICA generates three reports (some examples can be found in Appendixes E, and F). The first traces step-by-step the way Previous Stories are processed; it is very useful to study and analyse the way characters' contexts behave and therefore the way atoms are created in long-term memory. It also includes the graphic of the Tensional Representation of all Previous Stories. The second is a map of the Abstract Representation that includes all atoms and their position. Finally, the third is the familiar New Story Context's Report that traces the development of a new tale. For questions of space, only examples of the second and third report are presented in the thesis. However, notice that the first and third reports are the same; the only difference is that the former reports the Previous Stories and the latter the story in progress. The importance and utility of the last report is obvious. However, the other two reports are also indispensable to trace, analyse and understand many of the processes in MEXICA. For example, with the

help of the map of the Abstract Representation the user can figure out some of the consequences of modifying the ACAS-Constant.

### **E. Conclusion.**

One of the main purposes of a computer model is to allow easy testing of the hypothesis (represented by the model) under different circumstances, which in some cases can be difficult or not feasible to perform in human subjects. MEXICA has proved to be flexible enough to permit such testing under different circumstances. It is important to stress that such a flexibility is unique in MEXICA; that is, as far as the author of this work knows no other story-generator program gives the user this flexibility to experiment with the model it represents.

## **6.4 Summary.**

This chapter evaluated MEXICA as a computer model. It was divided in three sections.

The first section tested whether the engagement-reflection theory worked. The first step to show that the theory worked was to demonstrate that MEXICA embodied the theory's main postulates: production of material without the use of predefined story-structures or explicit goals, and the production of material as the result of interactions between engagement and reflection. The second step to evaluate the theory was to examine the outputs produced by the system. Criteria to perform this evaluation were established, and four stories developed by MEXICA assessed. To validate the results obtained from this evaluation and compare MEXICA's outputs against other programs stories, a questionnaire to discover people's opinion about computerised stories was designed and given to 50 subjects. The results from the questionnaire indicated that MEXICA ER2 obtained the highest score.

The second section analysed the creative process and the method used to produce interesting stories. Following Sharples (1996), creativity was defined as producing adequate and novel outputs. In this way, adequacy was assessed through the questionnaire in Section 6.1. Novelty was assessed by comparing the themes of the stories produced by MEXICA against the themes of the Previous Stories. During the evaluation MEXICA generated interesting stories by using degradation-improvement processes to create suspense. This method was analysed in Section 6.2; the results from the questionnaire showed that this was an appropriate method to produce interesting stories.

The last section examined all the facilities that MEXICA offered to experiment with the model. In MEXICA all knowledge in memory was built from the Primitive Actions and Previous Stories. These files were built by the user, allowing her/him to control the content of memory. Furthermore, a large number of parameters that constrain the story development process were modifiable by the user. Once the story was written, reports were printed to help analyse and interpret the results obtained from MEXICA. As far as the author of this work knows, no other system offers these facilities.